

الإدارة العامة للبحوث

العينات و تطبيقاتها في البحوث الاجتماعية

Frequency Percent Row Pct Col Pct	A	B	C	Total
15 - 19	3 7.69 30.00 21.43	5 12.82 50.00 31.25	2 5.13 20.00 22.22	10 25.64
20 - 24	7 17.95 46.67 50.00	6 15.38 40.00 37.50	2 5.13 13.33 22.22	15 38.46
25 - 29	0 0.00 0.00 0.00	2 5.13 40.00 12.50	3 7.69 60.00 33.33	
30 - 34	4 10.26	3 7.69	2 5.13	9 23.08

تأليف

د. عبدالرزاق أمين أبو شعر

ضو هيئة التدريب السابق بمعهد الإدارة العامة

بسم الله الرحمن الرحيم



الإدارة العـامـة للبحـوث

العينات
وتطبيقاتها فى البحوث الاجتماعية

تأليف

د. عبدالرزاق أمين أبو شعر

عضو هيئة التدريب السابق بمعهد الإدارة العامة

١٤١٨هـ / ١٩٩٧م

بطاقة الفهرسة

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

أبو شعر ، عبدالرزاق أمين

العينات وتطبيقاتها في البحوث الاجتماعية - الرياض .

٤٧٠ ص ؛ ١٢ × ١٨ سم

ردمك : ٠ - ٠٢١ - ١٤ - ٩٩٦٠

١ - العينات (إحصاء) ٢ - العنوان

ديوي ٥١٩,٥٢ ١٦/٢٤٨٦

رقم الإيداع : ١٦/٢٤٨٦

ردمك : ٠ - ٠٢١ - ١٤ - ٩٩٦٠

المحتويات

رقم الصفحة

المقدمة :	٧
الفصل الأول : أساسيات فى المعاينة	٩
١-١ تعاريف ومصطلحات أساسية	١٢
١-٢ أهم التوزيعات الاحتمالية	٢٥
١-٣ تقدير معالم المجتمع	٣٠
١-٤ أساليب جمع البيانات	٣٩
١-٥ أنواع العينات ومجال استخدامها	٤٦
الفصل الثانى : الخطوات الأساسية لتصميم العينة وجمع البيانات	٤٩
٢-١ خطوات تصميم العينة	٥١
٢-٢ إعداد الاستمارة الإحصائية	٦٧
٢-٣ البحث التجريبى	٧٣
٢-٤ جمع البيانات وتدقيقها	٧٣
٢-٥ مصادر الأخطاء فى العينات وكيفية التقليل منها	٧٥
الفصل الثالث : المعاينة العشوائية البسيطة	٨١
٣-١ تعريف المعاينة العشوائية البسيطة	٨٣
٣-٢ طرق اختيار العينة العشوائية البسيطة	٨٦
٣-٣ تقدير أهم معالم المجتمع	٨٩
٣-٤ تقدير حجم العينة	١١٠
الفصل الرابع : معاينة نسبة المجتمع	١١٧
٤-١ رموز وتعاريف	١١٩
٤-٢ تقدير نسبة المجتمع والقيمة الكلية للمجتمع	١١٩

رقم الصفحة

٤-٣ تباين التقديرات لمعاينة النسب وتقديراتها ١٢١
٤-٤ حدود الثقة لتقدير نسبة المجتمع ، وتقدير القيمة الكلية للمجتمع ١٣١
٤-٥ تحديد حجم العينة فى معاينة النسب ١٣٤

الفصل الخامس : المعاينة الطبقية العشوائية ١٣٩
٥-١ تعريف المعاينة الطبقية العشوائية ١٤١
٥-٢ رموز وتعاريف ١٤٢
٥-٣ خطوات اختيار المعاينة الطبقية العشوائية ١٤٦
٥-٤ تقدير معالم المجتمع باستخدام المعاينة الطبقية العشوائية ١٤٩
٥-٥ حدود الثقة لتقدير متوسط المجتمع ، وتقدير القيمة الكلية للمجتمع ١٧١
٥-٦ طرق تخصيص حجم العينة على الطبقات وتحديد حجم العينة ١٧٤
٥-٧ المقارنة بين المعاينة العشوائية البسيطة والمعاينة الطبقية العشوائية ١٩٥

الفصل السادس : المعاينة الطبقية للنسب ٢٠٥
٦-١ رموز وتعاريف ٢٠٧
٦-٢ تقدير نسبة المجتمع ٢٠٨
٦-٣ تباين التقديرات للمعاينة الطبقية للنسب وتقديراتها ٢١٠
٦-٤ حدود الثقة لتقديرات نسبة المجتمع ، والقيمة الكلية للمجتمع ٢١٥
٦-٥ تحديد حجم العينة فى المعاينة الطبقية للنسب ٢١٩

الفصل السابع : المعاينة المنتظمة ٢٢٧
٧-١ رموز وتعاريف ٢٢٩
٧-٢ طريقة اختيار العينة المنتظمة ٢٣٠
٧-٣ تقديرات أهم معالم المجتمع ٢٣٣
٧-٤ المقارنة بين المعاينة المنتظمة والمعاينات الأخرى وأشكال المجتمع ٢٤٢
٧-٥ حدود الثقة لتقديرات متوسط المجتمع ، والقيمة الكلية للمجتمع ٢٤٨

رقم الصفحة

٧-٦ تقدير نسبة المجتمع ٢٥٠
٧-٧ تحديد حجم العينة المنتظمة ٢٥٣
٧-٨ المعاينة الطبقية المنتظمة ٢٥٦
٧-٩ المعاينة المنتظمة المتكررة ٢٥٦

الفصل الثامن : المعاينة العنقودية البسيطة ٢٦٧
٨-١ تعريف المعاينة العنقودية البسيطة ٢٦٩
٨-٢ طريقة اختيار العينة العنقودية البسيطة ٢٧٠
٨-٣ رموز ومصطلحات ٢٧١
٨-٤ تقدير أهم معالم المجتمع ٢٧٣
٨-٥ حدود الثقة لتقدير متوسط المجتمع ، وتقدير القيمة الكلية للمجتمع ٢٧٩
٨-٦ تقديرات نسبة المجتمع ، وتباين نسبة المجتمع ٢٨٢
٨-٧ تحديد حجم العينة ٢٨٦

الفصل التاسع : المعاينة العنقودية ذات المرحلتين وذات المراحل المتعددة ٢٩٧
٩-١ تمهيد ٢٩٩
٩-٢ المعاينة العنقودية ذات المرحلتين ٢٩٩
٩-٣ المعاينة العنقودية ذات المراحل المتعددة ٣٣٢
٩-٤ المعاينة الطبقية العنقودية ٣٤٥
٩-٥ المعاينة العنقودية باحتمالات متناسبة مع الحجم ٣٤٦

الفصل العاشر : أنواع المعاينات الأخرى ٣٥١
١٠-١ المعاينة المزدوجة ٣٥٣
١٠-٢ المعاينة المتكررة فى مناسبات متعاقبة ٣٦٧
١٠-٣ المعاينة المساحية ٣٧٦
١٠-٤ المعاينة فى المجتمعات البرية ٣٧٧

رقم الصفحة

الفصل الحادى عشر : استخدام الحاسوب فى مجال العينات ٣٨٣

١١-١ تمهيد ٣٨٥

١١-٢ البرامج الإحصائية الجاهزة ٣٨٥

١١-٣ استخدام نظام (MINITAB) فى مجال العينات ٣٨٧

١١-٤ استخدام نظام ساس (SAS) فى مجال العينات ٣٩٣

الفصل الثانى عشر : حالة عملية عن استخدام العينات فى مجال البحوث ٤١١

١٢-١ مرحلة تصميم البحث ٤١٤

١٢-٢ مرحلة جمع البيانات ٤٢٦

١٢-٣ مرحلة تجهيز البيانات ٤٢٦

١٢-٤ مرحلة وصف وتحليل البيانات ٤٢٦

الملاحـــق ٤٤١

المراجـــع ٤٦٧



تطورت مختلف العلوم فى السنوات الأخيرة تطوراً كبيراً ، أدى إلى تحقيق الإنجازات التى نشاهدها فى مختلف المجالات الاقتصادية والاجتماعية . إن التطور السريع الذى تحقق فى مجالات الطب والزراعة والصناعة والفلك والفضاء والعلوم الإدارية والاقتصادية والعلوم الأخرى ، هو نتيجة للبحوث العلمية النظرية والتطبيقية التى قام بها الباحثون وتوصلوا فيها إلى النتائج الدقيقة التى أحدثت هذا التطور .

ويعدّ الإحصاء الأداة الرئيسية التى يستخدمها الباحثون لجمع البيانات المتعلقة ببحوثهم وتبويبها ووصفها وتحليلها ، وذلك للوصول إلى النتائج بشكل علمى وسليم . وقد أثبت أسلوب العينات ـ كأسلوب لجمع البيانات ـ نجاحاً كبيراً فى معاينة جزء من المجتمع الذى ندرسه للوصول إلى خصائص المجتمع الذى لا نستطيع دراسته بصورة شاملة لأسباب متعددة ، كضخامة الإمكانات المالية والبشرية التى يتطلبها الحصر الشامل إضافة للوقت الكبير الذى يستغرقه .

ولو أمعنا النظر فى مكتبتنا العربية ، لوجدنا نقصاً كبيراً فى الكتب التى تهتم بالأساس النظرى والخطوات العملية لجمع البيانات باستخدام أسلوب العينات . لذا فإننا نهدف من كتابنا هذا إلى تحقيق الأهداف التالية :

- توفير الأساس النظرى فى العينات لتمكين مستخدميها من اختيار نوع العينة المناسب ، وتحديد حجمها ، وتقدير المقاييس بدقة كبيرة .
- عرض بعض التطبيقات والحالات العملية التى توضح عملياً كيفية استخدام العينات فى المجالات العملية .
- توفير مرجع فى مجال العينات للباحثين والدارسين والمهتمين بالدراسات الاقتصادية والاجتماعية .

ولتحقيق هذه الأهداف ضَمَّنَّا الكتاب مقدمة واثنى عشر فصلاً ، إضافة للملاحق وقائمة المراجع . وفيما يأتى موجز عما تضمنته هذه الفصول من موضوعات اطلع القارئ على تفاصيلها فى ثبت المحتويات السابق :

- لقد تضمن الفصل الأول أهم المفاهيم والموضوعات الإحصائية التى تعد أساسيات فى المعاينة .
- وتضمن الفصل الثانى الخطوات الأساسية لتصميم العينة وجمع البيانات ميدانياً .
- وتضمنت الفصول الثمانية التالية (من الفصل الثالث إلى الفصل العاشر) الموضوعات المتعلقة بتعريف وطريقة اختيار العينة ، وتقدير المعالم ، وتحديد حجم العينة لكل نوع من أنواع العينات .
- وتضمن الفصل الحادى عشر الموضوعات التى توضح كيفية استخدام الحاسوب فى مجال العينات مع التركيز على نظام (MINITAB) ونظام ساس (SAS) .
- أما الفصل الثانى عشر ، فقد تضمن حالة عملية شاملة عن استخدام العينات فى مجال البحوث .

نأمل أن تتحقق الفائدة المرجوة من هذا الكتاب والله الموفق .

المؤلف

الفصل
الأول

أساسيات فى المعاينة

تمهيد :

يلعب الإحصاء دوراً بارزاً في المجالات الاقتصادية والاجتماعية ، حيث يهتم العاملون في هذه المجالات بالبيانات الإحصائية وتحليلاتها لاتخاذ القرارات السليمة المتعلقة برسم السياسات الاقتصادية والاجتماعية ومتابعة تنفيذها . ويعد الإحصاء الأداة الرئيسية التي يستخدمها المخططون في جميع الدول لإعداد خطط التنمية الاقتصادية والاجتماعية واتخاذ القرارات لتنفيذ هذه الخطط ، ومتابعة تنفيذها بالشكل المناسب .

ويمكننا تعريف الإحصاء بأنه «الأساليب والنظريات العلمية التي تهتم بجمع البيانات وعرضها ووصفها وتحليلها واستخدامها لأغراض اتخاذ القرارات أو التنبؤ أو التحقق من صحة نظرية معينة» .

إن تنفيذ البحث الإحصائي للحصول على البيانات التي يحتاجها المخططون والباحثون ، يتم على مراحل رئيسية تسمى **مراحل البحث الإحصائي** نلخصها بما يلي :

١ – **تصميم البحث** : تتضمن هذه المرحلة الخطوات المتعلقة بالأعمال التحضيرية التي تسبق عملية جمع البيانات ميدانياً . ويتم في هذه المرحلة تحديد المشكلة التي يعالجها البحث وأهدافه وموعد تنفيذه . كذلك يتم تحديد البيانات المطلوبة ووضع الفرضيات وطرق التحليل التي سيستخدمها الباحث لتصميم الاستمارة على ضوء هذه الخطوات . كما تتضمن هذه المرحلة تحديد أسلوب جمع البيانات المناسب وطريقة جمعها والخطوات الأخرى التي سندرسها في الفصل الثاني .

٢ – **جمع البيانات** : يتم في هذه المرحلة جمع البيانات ميدانياً حسب الخطة المحددة في مرحلة تصميم البحث .

٣ – **عرض البيانات** : يتم في هذه المرحلة تبويب البيانات يدوياً أو باستخدام الحاسوب (الحاسب الآلي) وذلك لعرضها في جداول أو رسوم بيانية .

٤ – **وصف البيانات** بمقاييس متعددة كمقاييس النزعة المركزية (المتوسطات) ومقاييس التشتت ومقاييس الالتواء والتفرطح .

٥ – **تحليل البيانات والنتائج** التي تم الوصول إليها لاتخاذ القرارات المناسبة ، أو التنبؤ بالقيم المستقبلية ، أو التحقق من صحة فرضيات ونظريات معينة .

٦ - اقتراح التوصيات المناسبة ونشر النتائج .

يلاحظ مما سبق ، أن إحدى الخطوات المهمة التي تتضمنها مرحلة تصميم البحث هي تحديد أسلوب جمع البيانات المناسب الذي سنستخدمه ، أي تحديد ما إذا كنا سنستخدم أسلوب الحصر الشامل أو أسلوب شبه الحصر أو أسلوب المعاينة .

لذلك فإننا سنقوم في كتابنا بدراسة كيفية تنفيذ البحث الإحصائي باستخدام أسلوب المعاينة الذي يسمى «البحث بالعينة» لتمييزه عن البحث الإحصائي باستخدام أسلوب الحصر الشامل .

وسنقوم بتوضيح أهم المفاهيم والمصطلحات اللازمة لدراسة الموضوعات المتعلقة بالبحوث التي تنفذ باستخدام أسلوب المعاينة .

١-١ تعاريف ومصطلحات أساسية .

١-١-١ وحدة المعاينة (Sampling Unit)

وحدة المعاينة هي «الجزء أو الكيان الصغير الذي نجمع منه البيانات» . إن كل وحدة من الوحدات المكونة للمجتمع هي وحدة معاينة أي أن عدد وحدات المعاينة هي عدد وحدات المجتمع . إن وحدات المعاينة قد تكون وحدات طبيعية تتعلق بالجنس البشري (كالموظف والطالب والفرد والأسرة) أو وحدات مصطنعة (كالمؤسسة أو الوزارة أو المسكن أو المصنع) . كما أن وحدات المعاينة قد تكون متشابهة من حيث الحجم أو مختلفة . وعند تنفيذ البحوث الميدانية ، يجب تحديد وتعريف وحدة المعاينة تعريفًا واضحًا لجمع البيانات من الوحدات التي يشملها البحث وعدم تداخل هذه الوحدات مع تلك التي لايشملها البحث .

كذلك يجب التمييز بين وحدات المعاينة ووحدات المشاهدة (وحدة المشاهدة هي الوحدة التي يجرى عليها القياس أو التصنيف) اللتين قد تتطابقان أو لا تتطابقان (مثلاً قد تكون وحدة المعاينة المصنع ووحدة القياس المدير أو العامل) .

١-١-٢ المجتمع الإحصائي (Statistical Population)

المجتمع الإحصائي هو عبارة عن «جميع وحدات المعاينة التي نقوم بدراستها» أي هو جميع وحدات المعاينة التي نريد الاستدلال على خواصه عن طريق العينة . ويمكننا تقسيم المجتمعات إلى مجتمعات ثابتة لاتخضع لتغيرات خلال فترة (قصيرة) من الزمن كالمدن والشوارع ، ومجتمعات غير ثابتة (حركية) تتغير بشكل سريع من فترة لأخرى مثل عدد

السكان وعدد السيارات التى تمر فى شارع ما . ويجب تحديد المجتمع الذى سيشمله البحث تحديداً واضحًا ودقيقًا لتعميم نتائج العينة بشكل دقيق ، خاصة فيما يتعلق بعدد وحدات المجتمع حيث يمكننا التمييز بين المجتمع المحدود (Finite Population) عندما يكون عدد القيم محدوداً والمجتمع غير المحدود (Infinite Population) عندما يتضمن المجتمع عدداً لا نهائيًا من القيم .

١-١-٣ العينة والمعاينة (Sample and Sampling)

نستخدم كلمة العينة كثيراً فى حياتنا اليومية ، إذ عندما يمرض شخص ما ، يطلب الطبيب فحص عينة من دمه أى بجزء منه . كذلك عندما نريد شراء سلعة معينه كالحبوب (القمح ، الأرز ، ..) نختار جزءًا من هذه السلعة للتأكد من جودتها ، ولاتخاذ قرار بشرائها أو عدم شرائها . إن عملية الاختيار قد تكون جيدة ومناسبة بحيث تمكننا من الوصول إلى القرار السليم ، وقد تكون خاطئة تعطى نتائج مضللة .

وتعرف العينة بأنها «جزء من المجتمع يتم اختياره لتمثيل المجتمع بأجمعه» . أما المعاينة فتعرف بأنها «عملية اختيار جزء من المجتمع الإحصائى للاستدلال على خواص المجتمع بأكمله عن طريق تعميم نتائج العينة » .

ولتوضيح هذين المفهومين ، نورد المثال التالى . نفرض أننا نريد دراسة مستوى الرضا الوظيفى لموظفى إحدى الجهات ، ونظراً لضخامة عدد موظفى هذه الجهة ، فقد تقرر اختيار عدد من الموظفين يمثلون المجتمع . إن الموظفين الذين تم اختيارهم هم العينة ، إذ يشكلون جزءًا من المجتمع يتضمن خصائصه . أما عملية اختيار هذه العينة وتعميم النتائج للاستدلال على خصائص المجتمع فتسمى «معاينة» .

١-١-٤ حجم المجتمع وحجم العينة (Population Size and Sample Size)

يقصد بحجم المجتمع عدد جميع وحدات المعاينة التى يتكون منها المجتمع ويرمز له عادة بالرمز (N) .

أما حجم العينة ، فهو عدد وحدات المعاينة التى تم اختيارها ويرمز له عادة بالرمز (n) . ويعتبر حجم العينة صغيراً إذا كان أقل من (٣٠) أى إذا كانت ($n<30$) .

١-١-٥ كسر المعاينة (Sampling Fraction)

يمثل كسر المعاينة نسبة الوحدات المختارة فى العينة إلى عدد وحدات المعاينـة فى المجتمع ، أى يساوى نسبة حجم العينة إلى حجم المجتمع ويرمز له عادة بالرمز (f) حيث $f = \frac{n}{N}$

وعندما يكون لدينا عينات جزئية (يشكل مجموعها العينة) أى $(n=n_1+n_2+\cdots+n_L)$ حيث (L) عدد الأقسام (كما هو الحال فى المعاينة الطبقية التى سندرسها فيما بعد) نجد أن كسر المعاينة للطبقة (i) يساوى $f_i = \frac{n_i}{N_i}$ ، حيث (N_i) حجم المجتمع فى الطبقة (i) و (n_i) حجم العينة فى الطبقة (i) ويكون لدينا عدة كسور للمعاينة (L كسراً) .

$$f_1 = \frac{n_1}{N_1} , f_2 = \frac{n_2}{N_2} , \ldots , f_L = \frac{n_L}{N_L}$$

١-١-٦ المتغير العشوائى (Random Variable)

عندما نقيس وزن أو طول أو عمر شخص ما ، فإننا نشير إلى النتائج التى نحصل عليها بقيم معينة يعبر عنها بمتغير . لذا يمكننا تعريف المتغير بأنه «رمز يمكن أن يأخذ أيّة قيمة سبق تحديدها تسمى مجال هذا المتغير» . ويرمز عادة للمتغيرات برموز (X, Y, Z,.......) . مثلا عندما يكون لدينا (N) قيمة تمثل أعمار الأطفال ، يمكننا التعبير عن هذه القيم بالمتغير (X) حيث : $X : X_1 , X_2 , \ldots , X_i , \ldots , X_N$

يشير الدليل إلى رقم القيمة أى رقم الوحدة الإحصائية وعندما نحصل على النتائج (القيم) نتيجة العوامل العشوائية (عوامل الحظ أو الصدفة (Chance Factors)) يسمى المتغير «متغير عشوائى» ، كما تسمى النتائج التى نحصل عليها بالمشاهدات أو المفردات (Observations) . ويمكننا تعريف المتغير العشوائى بأنه دالة ذات قيم عددية حقيقية معرفة على فضاء العينة .

ونستطيع التمييز بين نوعين من المتغيرات العشوائية :

أ – متغير عشوائى متقطع (Discrete Random Variable)

وهو المتغير العشوائى الذى نحصل عليه عندما يكون هناك تقطعات أو قفزات بين القيم ،

وعند عدم وجود قيم بين كل قيمتين من القيم ، ويأخذ عدداً محدوداً من القيم ، مثلاً عدد أفراد الأسرة للموظفين في إحدى الجهات هو متغير عشوائي متقطع يأخذ القيم :

$X : 0,1,2,3,4, ..., 10$

ب – متغير عشوائي متصل (Continuous Random Variable)

المتغير العشوائي المتصل هو المتغير العشوائي الذى لا يتضمن فجوات أو تقطعات كما هو الحال فى المتغير المتقطع . أى أن المتغير العشوائي المتصل هو المتغير الذى يمكن أن يأخذ أية قيمة ضمن مجال القيم للمتغير الذى ندرسه . مثلاً درجات حرارة المرضى (Y) يمكن أن تأخذ عدة قيم : تتراوح بين (٣٦) و (٤١) أى : $Y : 37,37.5, 39.1,38,38.2$.

إن البيانات التى يمكن التعبير عنها بمتغيرات متقطعة تسمى بيانات متقطعة ، والبيانات التى يمكن التعبير عنها بمتغيرات متصلة تسمى بيانات متصلة . وعندما يأخذ المتغير قيمة وحيدة فقط يسمى «ثابت» .

وكذلك يمكننا التمييز بين المتغيرات الكمية التى يمكن قياسها كالأطوال والأوزان وغيرها ، والمتغيرات النوعية أو الاسمية التى تعبر عن الظواهر التى لا يمكن قياسها كالجنس أو اللون ، مثلاً ، نعبر عن متغير الجنس للمرضى :

$X : 1,2,1,1,2.$

حيث يشير العدد (1) إلى المريض إذا كان ذكراً والعدد (2) يشير إليه إذا كان أنثى ، والجنس هو متغير اسمى .

١-١-٧ الوسط الحسابى (Arithmatic Mean)

يعد الوسط الحسابى أحد وأهم مقاييس النزعة المركزية ، ويعرف الوسط الحسابى بأنه القيمة التى نحصل عليها إذا قسمنا مجموع القيم على عددها . إذا رمزنا لقيم المجتمع بالمتغير (X) حيث لدينا (N) قيمة أو مفردة ، يكون الوسط الحسابى للمجتمع ولنرمز له بالرمز (μ) :

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \text{............ (1 - 1)}$$

حيث $\sum_{i=1}^{N} X_i$ يشير إلى مجموع قيم المجتمع التى عددها (N) قيمة .

وإذا رمزنا إلى قيمة العينة فى السحب (i) بـ (x_i) حيث لدينا i = 1,2,...,n ، فإن الوسط الحسابى للعينة ولنرمز له بالرمز ($\bar{x}$) وتقرأ (x bar) يساوى

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad .... (1 - 2)$$

ويعد متوسط العينة من أفضل المقدرات لمتوسط المجتمع الذى يكون غالبًا غير معلوم ، لأن قيم المجتمع غير معلومة فى معظم الحالات .

وكثيرًا ما تستخدم كلمة المتوسط (MEAN) للدلالة على الوسط الحسابى .

١-١-٨ التباين والانحراف المعيارى (Variance and Standard Deviation)

يعد التباين والانحراف المعيارى من أهم مقاييس الانتشار أو التشتت (Measures of Variation or Dispersion) التى تقيس مدى انتشار القيم عن بعضها أو عن قيمة معينة . ويعد التباين أحد المقاييس التى تستخدم لقياس مدى ابتعاد القيم عن الوسط الحسابى ، إذ كلما كانت القيم بعيدة عنه كان التباين أكبر ، والتباين هو عبارة عن مجموع مربعات انحرافات القيم عن وسطها الحسابى مقسومًا على عددها ، ويمكننا التمييز بين تباين المجتمع (σ^2) وتباين العينة (s^2) .

– تباين المجتمع (Population Variance) ويساوى :

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N} \qquad .... (1 - 3)$$

حيث (μ) هو الوسط الحسابى للمجتمع و (N) حجم المجتمع .

- تباين العينة (Sample Variance) ويساوى :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1} \quad \text{............ (1 - 4)}$$

حيث $(\bar{x})$ هو الوسط الحسابى للعينة و (n) حجم العينة . وعندما يـكون حجم العينة $(n \geq 30)$ نضع فى المقام (n) عوضـًا عن (n-1) . أما الانحراف المعيارى فـهو عبارة عن الجذر التربيعى للتباين ويكون لدينا :

- الانحراف المعيارى للمجتمع (σ) ويساوى :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}} \quad \text{............ (1 - 5)}$$

- الانحراف المعيارى للعينة (s) ويساوى :

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}} \quad \text{............ (1 - 6)}$$

وكثيرًا ما نستخدم الصيغة التالية لحساب الإنحراف المعيارى للعينة :

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)} \quad \text{............ (1 - 7)}$$

كما يرمز أحيانًا لتباين المجتمع بالرمز V (X) أو VAR (X) ولتباين العينة $V(x)$ أو $VAR(x)$ ويكون الانحراف المعيارى للعينة :

$$s = \sqrt{VAR(x)}$$

١-١-٩ التغاير والارتباط (Covariance and Correlation)

نفترض أننا نرغب فى دراسة العلاقة بين متغيرين عشوائيين (X) و (Y) لعينة حجمها (n) وحدة ، فيكون لدينا (n) زوج من القيم $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

إن متوسط مجموع حاصل ضرب انحرافات القيم عن الوسط الحسابى للمتغيرين هو التغاير أى أن التغاير ولنرمز له بالرمز $COV(x, y)$ يساوى :

$$COV(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \qquad \ldots\ldots\ldots\ldots (1-8)$$

ونضع (n) عوضاً عن (n-1) إذا كان حجم العينة كبيراً $(n \geq 30)$

ويستخدم التغاير كمقياس نوعي لمدى وجود علاقة بين المتغيرين (X) و(Y) . عندما يكون التغاير مساوياً للصفر ، يعنى ذلك عدم وجود علاقة بين المتغيرين .

ومن الصعب استخدام التغاير كمقياس لدرجة قوة العلاقة بين المتغيرين لأن قيمته تعتمد على نوع المقياس المستخدم ، لذا من الصعب تحديد ما إذا كان التغاير كبيراً من نظرة سريعة ، لذا يستخدم معامل الارتباط كمقياس كمى لقياس درجة قوة العلاقة بين متغيرين ، وكثيراً ما تستخدم الصيغة التالية لاستخراج التغاير بين متغيرين :

$$COV(x, y) = \frac{\sum_{i=1}^{n} x_i \, y_i - n\,\bar{x}\,\bar{y}}{n-1} \qquad \ldots\ldots\ldots\ldots (1-9)$$

ونستخدم معامل الارتباط لقياس درجة قوة الارتباط الخطى بين متغيرين ولنرمز له بالرمز (r) ويساوى :

$$r = \frac{COV(x, y)}{s_x \, s_y} \qquad \ldots\ldots\ldots\ldots (1-10)$$

حيث (s_x) و (s_y) هما الانحراف المعيارى للمتغيرين (X) و(Y) على التوالى . وتتراوح قيمة معامل الارتباط بين 1- و 1+ أى : $-1 \le r \le +1$ حيث يساوى (1-) عندما يكون الارتباط بين المتغيرين (X) و (Y) تامًا وسالبًا ، ويساوى هذا المعامل (1+) عندما يكون الارتباط بين هذين المتغيرين تامًا وموجبًا ، ويساوى الصفر عندما يكون الارتباط الخطى البسيط معدومًا ، أى لا يوجد ارتباط بين المتغيرين . ويمكننا استخدام إحدى الصيغتين التاليتين لحساب معامل الارتباط :

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\, s_x\, s_y} \qquad \text{............ (1 - 11)}$$

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}}{(n-1)\, s_x\, s_y} \qquad \text{............ (1 - 12)}$$

حيث

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}{n-1}}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - n \bar{y}^2}{n-1}}$$

إن الصيغ السابقة لاستخراج معامل التغاير ومعامل الارتباط من بيانات العينة هى مقدرات للمعالم المقابلة لها فى المجتمع .

تطبيق (١-١)

اختيرت عينة عشوائية حجمها (٥) أشخاص لدراسة مدى وجود علاقة بين دخلهم (x) وإنفاقهم (y). وكانت بيانات الدخل والإنفاق الشهرى (بآلاف الريالات) كما يلي :

x : 4, 6, 7, 5, 3

y : 3, 5, 5, 4, 3

المطلوب استخراج :

- الوسط الحسابى للدخل والإنفاق الشهرى .
- التباين والانحراف المعيارى للدخل والإنفاق .
- التغاير بين المتغيرين (x) و (y) .
- معامل الارتباط بين الدخل والإنفاق .

الحـــــل :

من بيانات التطبيق نجد أن

$$\sum_{i=1}^{n} x_i = 25 \quad \sum_{i=1}^{n} y_i = 20 \quad n = 5 \quad \sum_{i=1}^{n} x_i^2 = 135 \quad \sum_{i=1}^{n} y_i^2 = 84$$

$$\sum_{i=1}^{n} x_i \; y_i = 106$$

- الوسط الحسابى للدخل $(\bar{x})$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{5} (4 + 6 + ... + 3) = \frac{25}{5} = 5$$

أى (٥٠٠٠) ريال .

- الوسط الحسابى للإنفاق ($\overline{y}$) :

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{5}(20) = 4$$

أى (٤٠٠٠) ريال ·

- التباين والانحراف المعيارى

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right)$$

ويكون التباين للمتغير (x) :

$$s_x^2 = \frac{1}{5-1}(135 - 5 \times 5^2) = 2.5$$

والانحراف المعيارى للمتغير (x) :

$$s_x = \sqrt{s_x^2} = \sqrt{2.5} = 1.5811$$

أما التباين للمتغير (y) يساوى :

$$s_y^2 = \frac{1}{5-1}(84 - 5 \times 4^2) = 1$$

والانحراف المعيارى للمتغير (y) يساوى (1) .

- التغاير COV (x,y) :

$$\mathrm{Cov}(x, y) = \frac{\sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y}}{n-1} = \frac{106 - 5 \times 5 \times 4}{5-1} = 1.5$$

- معامل الارتباط للمتغيرين (x, y)

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y}}{(n-1)\,s_x\,s_y} = \frac{106 - 5 \times 5 \times 4}{4 \times 1.5811 \times 1} = 0.949$$

أى أن هناك ارتباطًا خطيًا موجبًا (طرديًا) قويًا للغاية بين الدخل والإنفاق .

١-١-١٠ معلمة المجتمع (A Population Parameter)

عند دراسة متغير عشوائى (X) فإن دالة كثافة احتمالة تعتمد عادة على مقياس أو عدة مقاييس (ثوابت) كالوسط الحسابي والتباين ٠ إن معرفة هذه المقاييس تحدد الخصائص الأساسية للمتغير موضوع الدراسة وتسمى الثوابت التى تعتمد عليها دالة كثافة الاحتمال معالم المجتمع ٠

إن معلمة المجتمع تعبير عددى يلخص خصائص جميع قيم المجتمع إذا كانت غير خاضعة للأخطاء ٠ ويتم حساب معالم المجتمع عند استخدام أسلوب الحصر الشامل بشكل تام ودقيق أى عندما لاتقع أخطاء ٠ ويعد الوسط الحسابى للمجتمع (μ) وتباينه (σ^2) من أهم معالم المجتمع حيث :

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$

١-١-١١ إحصائية العينة (A Sample Statistic)

غالبًا ما تكون معالم المجتمع مجهولة حيث نقوم بتقديرها من بيانات عينة تمثل المجتمع ٠ إن إحصائية العينة هى مقدر لمعلمة المجتمع يتم حسابها من بيانات العينة التى تمثل هذا المجتمع ٠ ويعد الوسط الحسابى للعينه ($\overline{x}$) وتباين العينة (s^2) y من إحصائيات العينة حيث :

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$s_i^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

١-١-١٢ الاحتمال (Probability)

كثيراً ما نستخدم مفهوم الاحتمال فى حياتنا اليومية كأن نقول إن احتمال نجاح الطالب فى مادة الرياضيات ٥٠٪ أو ٧٠٪ (٠,٥٠ أو ٠,٧٠) ٠

ويتراوح الاحتمال بين الصفر والواحد ، إذ كلما كان الحدث أكثر وقوعاً كان الاحتمال أقرب إلى الواحد ، وكلما كان الحدث أقل وقوعاً كان الاحتمال أقرب إلى الصفر. إن احتمال وقوع الحدث الأكيد يساوى الواحد واحتمال عدم وقوعه إطلاقاً يساوى الصفر. لنرمز إلى احتمال حدوث الحدث (E) بالرمز p (E) واحتمال عدم حدوث الحدث (E) بالرمز q (E) حيث :

$$q(E) = 1 - p(E)$$

وتستخدم كلمة «نجاح» للإشارة إلى وقوع الحدث وكلمة «فشل» لعدم وقوعه . وللوصول إلى تعريف دقيق للاحتمال ، لابد لنا من تعريف التجربة والحدث . وتعرف التجربة (An Experiment) بأنها «عملية تجرى تحت ظروف معينة ولايمكن التنبؤ بنتيجتها بشكل أكيد» وللتجربة نتائج محتملة (Possible Outcomes)) . أما الحدث (An Event) فهو مجموعة النتائج التى لها خصائص محددة فى المجموعة الكلية للنتائج (Ω) .

إذا رمزنا إلى عدد النتائج المحتملة بـ N (Ω) وعدد النتائج (الحالات) المواتية (التى نحصل عليها نتيجة الحدث E) بـ n (E) يكون احتمال حدوث الحدث (E) ولنرمز له بالرمز P (E) مساوياً لعدد الحالات المواتية مقسوماً على عدد الحالات الممكنة ، وذلك عندما يكون لجميع النتائج الممكنة فى (Ω) الفرصة نفسها فى الحدوث ، أى أن :

$$P(E) = \frac{n(E)}{N(\Omega)} = \frac{n}{N} \qquad \text{............ (1 - 13)}$$

ولتوضيح المفاهيم السابقة ، نورد المثال الآتى . إذا أردنا استخراج احتمال اختيار موظف لديه شهادة الماجستير من موظفى إحدى الجهات البالغ عددهم (٢٠٠) موظف إذا كان عدد الذين لديهم ماجستير فى هذه الجهة هو (١٠) موظفين ، فنجد من هذا المثال أن التجربة هى اختيار الموظف للتعرف على مؤهله ، ولدينا عدة حوادث .... E_1 , E_2 , E_3 , E_4 , حسب مؤهلات الموظفين حيث E_1 ترمز للحدث إذا كان الموظف الذى تم اختياره يحمل مؤهل الماجستير و E_2 إذا كان مؤهله البكالوريوس وهكذا ، ويكون عدد الحالات الممكنة (N = 200) وعدد الحالات المواتية (n = 10) وبالتالى يكون احتمال الحصول على موظف اختير عشوائياً ومؤهله ماجستير $p(E_1)$ يساوى :

$$P(E_1) = \frac{n(E_1)}{N(\Omega)} = \frac{n}{N}$$

$$= \frac{10}{200} = 0.05$$

أى ٥٪ . أما احتمال اختيار موظف مؤهله ليس بشهادة ماجستير فيساوى

$$q(E_1) = 1 - p(E_1) =$$

$$= 1 - 0.05 = 0.95$$

أى يساوى ٩٥٪ .

١-١-١٢ التوقع (Expectation)

إذا كان لدينا متغير عشوائى (X) يمثل عدد أفراد الأسرة لموظفى إحدى الإدارات حيث

$$X : x_1, x_2, \ldots, x_n$$

وكانت دالة احتمال أن يكون عدد أفراد الأسرة (x_i) هو $f(x_i)$ فإن التوقع (ويسمى أحياناً التوقع الرياضى أو القيمة المتوقعة) ولنرمز له بالرمز $E(X)$ يساوى :

$$E(X) = \sum_{i=1}^{n} f(x_i)\, x_i \qquad \ldots\ldots\ldots\ldots (1 - 14)$$

وتوجد صيغة أخرى للتوقع إذا كان المتغير العشوائى متصلاً باستخدام التكامل :

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx \qquad \ldots\ldots\ldots\ldots (1 - 15)$$

حيث $f(x)$ هى دالة كثافة الاحتمال للمتغير العشوائى (X) .

إن القيمة المتوقعة هى الوسط الحسابى للمجتمع (μ) لأننا إذا استبدلنا فى صيغة التوقع $f(x_i)$ بالتكرارات النسبية (f_i/n) حيث ($n = \Sigma f_i$) ، فإن التوقع يصبح ($\Sigma f_i x_i/n$) أى الوسط الحسابى للعينة التى حجمها (n) .إن التكرارات النسبية (f_i/n) تقترب من الاحتمالات $f(x_i)$ كلما زادت قيمة (n) ويؤدى ذلك إلى تفسير $E(X)$ كقيمة تمثل متوسط المجتمع الذى سحبت منه العينة *

* لمزيد من التفاصيل ، راجع :
سبيجل موراى : ملخصات شوم ، نظريات ومسائل فى الإحصاء ، ترجمة شعبان عبدالحميد شعبان ، دار ماكجروهيل للنشر ، ١٩٧٨م ، ص ١٦١ .

تطبيق (١-٢) :

فيما يأتى توزيع موظفى إحدى الجهات حسب عدد أفراد أسرهم والاحتمالات المقابلة لحجم الأسرة للموظف :

عدد أفراد الأسرة (X)	صفر	١	٢	٣	٤	٥
الاحتمال $f(X)$	٠,٠٥	٠,٠٥	٠,٢٠	٠,٣٥	٠,٣٠	٠,٠٥

ما هو متوسط عدد أفراد الأسرة للموظفين (أى القيمة المتوقعة) .

الحـــل :

$$E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

$$= (0x0.05) + (1x0.05) + (2x0.20) + (3x0.35) + (4x0.30) + (5x0.05)$$
$$= 2.95 \approx 3$$

أى متوسط عدد أفراد الأسرة لمجتمع الموظفين تقريباً ثلاثة أفراد .

١-٢ أهم التوزيعات الاحتمالية .

عند دراستنا للتوزيعات الاحتمالية ، نميز بين نوعين من هذه التوزيعات :

- التوزيعات الاحتمالية للمتغيرات المتقطعة .
- التوزيعات الاحتمالية للمتغيرات المتصلة .

ويعد توزيع ذى الحدين من أهم التوزيعات الاحتمالية للمتغيرات المتقطعة ، كما يعد التوزيع الطبيعى وتوزيع ستيودنت (ت) من أهم التوزيعات للمتغيرات المتصلة . ولهذه التوزيعات أهمية خاصة عند دراسة العينات لاستخدامها عند تقدير معالم المجتمع . وسنقوم بدراسة هذه التوزيعات الاحتمالية باختصار .

١-٢-١ توزيع ذى الحدين (Binomial Distribution)

عندما نجرى تجربة ما ، فإنه عندما يقع الحدث ، نستخدم كلمة نجاح (كلمة نجاح تستخدم للإشارة إلى وقوع الحدث) ، وعندما لا يقع الحدث نستخدم كلمة فشل . وعندما نجرى

التجربة (n) مرة نستخدم متغيراً عشوائياً (X) يمثل العدد الكلي لمرات النجاح التى حصلنا عليها أى عدد مرات وقوع الحدث (النجاح) عند تكرار التجربة (n) مرة . ويسمى المتغير الذى من هذا النوع متغير بحدين .

وعندما نقوم بإعداد جدول يحتوى على المتغير العشوائى (X) والاحتمالات المقابلة لكل قيمة (x_i) f ، نحصل على ما يسمى جدول توزيع المتغير العشوائى .

إن الصيغة المستخدمة لحساب الاحتمالات للقيم الممكنة للمتغير العشوائى ، والتى تسمى دالة الاحتمال لتوزيع ذى الحدين ، ولنرمز له بالرمز f (x) ، وذلك عندما تكون نتائج التجربة فى المحاولات المختلفة مستقلة عن بعضها البعض ونجد أن :

$$f(x) = \frac{n!}{x!\,(n-x)!}\, p^{x}\, q^{n-x} \qquad \text{............ (1 - 16)}$$

حيث :

p احتمال حدوث الحدث فى المحاولة الواحدة للتجربة

q احتمال عدم حدوثه حيث $p + q = 1$

n عدد مرات تكرار التجربة

x عدد مرات النجاح التى سنحصل عليها

n! تقرأ مضروب (n) وهى عبارة عن حاصل ضرب كل الأعداد الصحيحة من (١) إلى (n) مثلاً !4 تساوى 4x3x2x1 كما أن مضروب الصفر !0 يساوى الواحد .

وعند استخدام توزيع ذى الحدين ، فإن الوسط الحسابى لمتغير ذى الحدين يساوى

$$\mu = n\,p \qquad \text{............ (1 - 17)}$$

وتباينه يساوى

$$\sigma^2 = n\,pq \qquad \text{............ (1 - 18)}$$

وسنستخدم هذا التوزيع فى الفصول القادمة عند دراسة تقدير نسبة المجتمع .

١-٢-٢ التوزيع الطبيعى (Normal Distribution)

يعد التوزيع الطبيعى (أو المعتاد) أحد الأمثلة المهمة للتوزيع الاحتمالى للمتغير المتصل . ويستخدم هذا التوزيع كثيراً فى مجال العينات . ويتصف هذا التوزيع بعدة خصائص :

- المتغير العشوائى المتصل (X) يأخذ قيماً من $-\infty$ إلى $+\infty$.
- أن شكل منحنى التوزيع الطبيعى يشبه الجرس .
- أن قمة المنحنى تقع عند متوسط المجتمع (μ) والمنحنى متماثل حول (μ) إذ كل طرف هو صورة مطابقة للطرف الآخر .
- يعتمد التوزيع الطبيعى على معلمتين هما متوسط المجتمع (μ) وتباين المجتمع (σ^2) لذا يشار إلى هذا التوزيع بالرمز $N(\mu, \sigma^2)$ حيث تشير (N) إلى (Normal) .
- أن مركز التوزيع يعتمد على (μ) وشكله يعتمد على الانحراف المعيارى (σ) .
- أن دالة كثافة الاحتمال للتوزيع الطبيعى هى :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 (x-\mu)^2 / \sigma^2} \qquad \text{............ (1 - 19)}$$

حيث : $-\infty < X < \infty$

μ الوسط الحسابى للمجتمع

σ الانحراف المعيارى للمجتمع

e قيمة ثابتة تساوى تقريباً $e = 2.71828$

π قيمة ثابتة تساوى تقريباً $\pi = 3.14159$

$f(x)$ دالة كثافة الاحتمال

وهناك ما يسمى التوزيع الطبيعى المعيارى (Standard Normal Distribution)

وهو توزيع طبيعى متوسطه $\mu = 0$ وتباينه $\sigma^2 = 1$ ويرمز لهذا التوزيع $N(0,1)$ ويستخدم الرمز (Z) للإشارة إلى المتغير المعيارى العشوائى الذى له توزيع طبيعى . ويتم حساب احتمالات أى متغير له توزيع طبيعى من احتمالات منحنى التوزيع الطبيعى المعيارى وفقاً للصيغة الآتية :

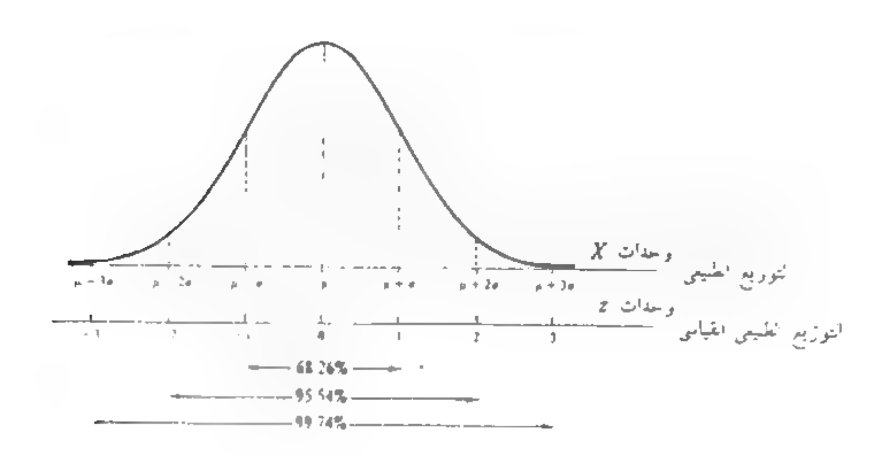
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 Z^2} \qquad \text{............ (1 - 20)}$$

حيث

$$z = \frac{x - \mu}{\sigma}$$

والشكل الآتــى يوضح المنحنــى الطبيعى المعيـارى ، حيث نلاحظ أن المساحة الواقعة بين ($z = -1 , +1$) هى (68.27%) وبين ($z = -2 , +2$) هى (95.45%) وكذلك بين ($z = -3 , +3$) هى (99.73%) وذلك من المساحة الكلية التى تساوى واحداً .

وقد تم إعداد جداول توضح المساحة تحت المنحنى المحصورة بين الإحداثى ($z = 0$) وأية قيمة موجبة لـ (z) ، ومن هذا الجدول فإن المساحة بين أية نقطتين يمكن حسابها باستخدام تماثل المنحنى حول ($z=0$) كما هو موضح فى الملحق رقم (1) .



شكل رقم (١)

منحنى التوزيع الطبيعى

١-٢-٢ توزيع ستيودنت (t) (Student Distribution)

يستخدم التوزيع الطبيعي للاستدلال على متوسط المجتمع عندما يكون تباين المجتمع (σ^2) معلومًا ، أو تكون العينة كبيرة بشكل كاف ، لنتمكن من الاستعاضة عن هذا التباين بتقديره من العينة (s^2) . ولكن عندما يكون تباين المجتمع غير معلوم وحجم العينة صغيراً (تكون العينة صغيرة إذا كان حجمها أقل من (٣٠) أى عندما يكون ($n < 30$)) ، نستخدم متغيراً جديداً يسمى متغير توزيع (t) أو ستيودنت وصيغته :

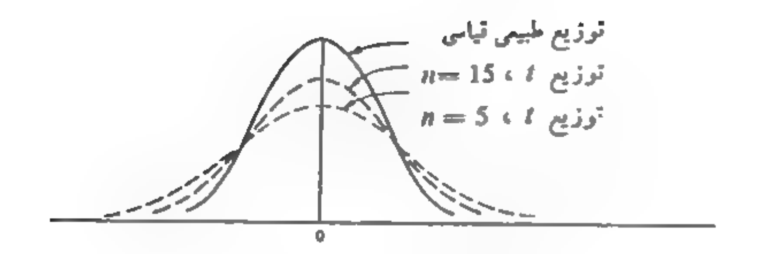
$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \qquad \text{............ (1 - 21)}$$

ويشبه هذا المتغير الطبيعي المعيارى (Z) باستثناء القيم الصغيرة جداً للعدد (n) وتختلف عنه في استخدامنا الانحراف المعيارى للعينة (s) ، وهذه ميزة تساعدنا على تقدير معالم المجتمع ، خاصة إذا كان حجم العينة صغيراً .

وعند اختيار عدد كبير من العينات ، حجم كل منها (n) وحدة من مجتمع طبيعى ، نحصل على عدد كبير من قيم (t) ، ويمكننا الحصول على التوزيع الاحتمالى لـ (t) والذى دالة كثافة احتماله :

$$f(x) = \frac{Y_o}{\left(1 + \frac{t^2}{n - 1}\right)^{n/2}} \qquad \text{............ (1 - 22)}$$

حيث (Y_o) مقدار ثابت يجعل المساحة تحت المنحنى مساوية للواحد و (n - 1) هو عدد درجات الحرية . إن المتغير (t) يتبع توزيع (t) إذا كان توزيع المجتمع طبيعاً . كذلك نجد أن هذا التوزيع يكون قريباً جداً من التوزيع الطبيعى عندما يكون حجم العينة كبيراً . وقد تم إعداد جداول توزيع (t) توضح الاحتمالات لقيمة (t) بمستويات متعددة ودرجات حرية (n - 1) متعددة أيضاً (ملحق رقم (2)) ويوضح الشكل الآتى منحنى توزيع ستيودنت لدرجات حرية (5) و (15) حيث يلاحظ اقترابه من منحنى التوزيع الطبيعى بازدياد حجم العينة (n) .



شكل رقم (٢)

التوزيع الطبيعي ومنحنى توزيع (t)

١-٢ تقدير معالم المجتمع (Estimation of Population Parameters)

عندما نقوم بدراسة ظاهرة معينه من بيانات المجتمع نحصل على معلمتي المجتمع (μ) و (σ^2) ، ولكن في كثير من الحالات ، نجد أن هاتين المعلمتين غالبًا ما تكونان مجهولتين ، فنقوم بتقديرهما من بيانات عينة يتم اختيارها عشوائيًا لتمثيل المجتمع تمثيلا حقيقيًا . وسنقوم بدراسة أهم الموضوعات المتعلقة بتقدير معالم المجتمع للتمييز بين مفهومي التقدير والمقدر وخواص المقدر الجيد وأنواع التقدير .

١-٢-١ التقدير والمقدر (Estimate and Estimator)

عندما نسحب عينة ما مفرداتها $x_1, x_2, \dots, x_n$ ونقوم بتقدير ثوابت دالة كثافة الاحتمال باستخدام هذه المفردات ، فإن القيمة المقدرة لكل ثابت تسمى تقديرًا .

أما الصيغة التي تستخدم للوصول إلى التقدير ، فتسمى مقدرًا وهو عبارة عن الدالة التي تعتمد على المفردات ، بينما التقدير عبارة عن قيمة الدالة عند وضع قيم المشاهدات فيها .

إن قيمة متوسط العينة ($\bar{x}$) هو تقدير المتوسط للمجتمع (معلمة المجتمع) أي ($\hat{\mu} = \bar{x}$) . أما الدالة المستخدمة لتقدير المتوسط فهي عبارة عن المقدر أي يساوي

$$\bar{x} = \hat{\mu} = f(x_1, x_2, \dots, x_n)$$

وبصيغة أخرى نجد أن المقدر يساوي

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

١-٣-٢ المقدر الجيد

إن المقدر لا يختلف من عينة لأخرى إلا إذا تغيرت صيغة هذا المقدر ، بينما يختلف التقدير من عينة لأخرى عند استخدام المقدر نفسه لاختلاف قيم المفردات من عينة لأخرى . وقد تكون القيمة المقدرة قريبة جداً من القيمة الحقيقية للمجتمع أو بعيدة عنها .

ويعد المقدر جيداً إذا كان في المتوسط لعدد كبير من العينات يعطى قيماً قريبة جداً من القيم الحقيقية للمجتمع ، والمقدر الأقرب إلى معلمة المجتمع هو المقدر الأفضل . وتوجد عدة خواص للمقدر الجيد تساعدنا على استخدامه لتقدير معالم المجتمع عندما تكون مجهولة .

١-٣-٢ خواص المقدر الجيد

للمقارنة بين المقدرات المختلفة ، توجد خواص معينة عندما تتحقق في المقدر يعد محققاً لصفات الجودة ، وهذه الخواص هي :

- عدم التحيز .
- الاتساق .
- الكفاءة .
- الكفاية .

ونظراً لأهمية هذه الخواص عند دراسة موضوعات المعاينة ، سنقوم بتعريفها باختصار .

١ – عدم التحيز Unbiasedness

يسمى المقدر ($\hat{\theta}$) مقدراً غير متحيز للمعلمة (θ) إذا كان توقعه يساوى هذه المعلمة أى عندما :

$$\boxed{E(\hat{\theta}) = \theta} \qquad \text{..........} (1 - 23)$$

وذلك لجميع قيم θ فى Ω_θ حيث تتضمن (Ω_θ) جميع قيم (θ)

أمثلـــــة :

– الوسط الحسابى لعينة عشوائية سحبت من مجتمع متغيره العشوائى (X) وتوقعه (μ) يساوى :

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

هو مقدر غير متحيز لـ (μ) وذلك لأن

$$E(\bar{x}) = \mu$$

- تباين عينة عشوائية مسحوبة من مجتمع إحصائى متغيره العشوائى X وتباينه σ^2 يساوى :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

يعد مقدراً غير متحيز لتباين المجتمع S^2 وذلك لأن :

$$E(s^2) = \sigma^2 \frac{N}{N-1} = S^2$$

٢ - الاتساق (Consistency)

إذا كان $(\hat{\theta})$ مقدراً للمعلمة (θ) محسوباً من مفردات عينة حجمها (n) فإن معنى الاتساق أن يؤول المقدر $(\hat{\theta})$ احتمالياً إلى القيمة الحقيقية للمعلمة (θ) عندما يزداد حجم العينة ويصبح قريباً من اللانهاية أى أن :

$$\lim_{n \to \infty} P\left[\left|\hat{\theta} - \theta\right| > \varepsilon\right] = 0 \qquad \text{............ (1 - 24)}$$

عندما $\varepsilon > 0$

ويتم ذلك عندما يتحقق الشرطان الآتيان :

$$\lim_{n \to \infty} E(\hat{\theta}) \longrightarrow \theta$$

$$\lim_{n \to \infty} V(\hat{\theta}) \longrightarrow 0$$

تطبيق (١ - ٢) :

سحبت عينة عشوائية عدد مفرداتها (n) من مجتمع متوسطه (μ) وتباينه (σ^2) . أن متوسط العينة $(\bar{x})$ مقدر متسق للمعلمة (μ) . لإثبات ذلك نعلم أن :

$$\lim_{n \to \infty} E(\bar{x}) \longrightarrow \mu$$

$$\lim_{n \to \infty} V(\bar{x}) = \lim_{n \to \infty} (\sigma^2/n) \longrightarrow 0$$

أى تحقق الشرطان اللازمان لاعتبار $(\bar{x})$ مقدراً متسقاً .

٣ - الكفاءة (Efficiency)

إذا كان لدينا مقدران غير متحيزين للمعلمة (θ) هما (θ_1 , θ_2) وكان تباين المقدر الأول أصغر من تباين المقدر الثانى أى

$$V(\theta_1) < V(\theta_2)$$

يعد المقدر (θ_1) أكفأ من المقدر (θ_2) .

وعادة عندما يكون لدينا مقدران ، نفضل المقدر الذى يكون متمركزاً حول المعلمة .

تطبيق (١-٤)

لمقارنة مدى تمركز الوسط الحسابى $(\bar{x})$ مع مدى تمركز الوسيط (ME) (المقدران غير متحيزين) ، نلجأ إلى مقارنة تباين المقدرين ونختار المقدر ذا التباين الأصغر .

إن تباين الوسط الحسابى والوسيط للعينات الكبيرة هما :

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

$$V(ME) = \frac{\pi \sigma^2}{2n}$$

حيث $\pi = 3.1416$

إذا كان حجم العينة محدداً فإن

$$\frac{V(\bar{x})}{V(ME)} = \frac{2}{\pi} = 0.636$$

وهذا يعنى أن تباين الوسط الحسابى أصغر من تباين الوسيط وبالتالى يكون المقدر $(\bar{x})$ أكفأ من المقدر (ME) .

٤ – الكفاية (Sufficiency)

يسمى المقدر $(\hat{\theta})$ مقدراً كافياً للمعلمة (θ) إذا كان الاحتمال الشرطي الحصول على العينة المستخدمة في التقدير إذا علم المقدر $(\hat{\theta})$ خالياً من المعلمة الحقيقية (θ). وبمعنى آخر نجد أن المقدر $(\hat{\theta})$ قد امتص جميع المعلومات المتوافرة عن المعلمة (θ) مجهولة القيمة ، بحيث بعد معرفة $(\hat{\theta})$ نجد أن المعلومات المتبقية لا تفيد في معرفة (θ) . ويمكننا إثبات كفاية المقدر $(\hat{\theta})$ باستخدام طريقة التحليل العاملي .

وإذا سحبنا عينة عشوائية حجمها (n) مفردة $(x_1, x_2, ..., x_n)$ وكانت دالة كثافة احتمال كل من هذه المفردات متشابهة $f(x, \theta)$ فإن دالة كثافة الاحتمال المشتركة لهذه القيم العشوائية تساوى :

$$g(x_1, x_2, .., x_n; \theta) = f(x_1, \theta) \quad f(x_2, \theta) .. ..f(x_n, \theta)$$

فإذا استطعنا صياغة هذه الدالة بالشكل :

$$g(x_1, x_2, .., x_n; \theta) h(\hat{\theta}, \theta) = k(x_1, x_2, .., x_n)$$

حيث $K(x_1, x_2, ..., x_n)$ دالة لا تحتوى على المعلمة (θ) فإننا نسمي $(\hat{\theta})$ بأنه مقدر كاف للمعلمة (θ) .

تطبيق (١–٥) :

إذا كان الوسط الحسابي للعينة $(\bar{x})$ هو مقدر لتوقع المجتمع (a) فيمكن إثبات أن هـذا المقـدر كـاف لتوقع المجتمع المعتاد (الطبيعي) μ .

نعلم أن احتمال الحصول على هذه العينة (دالة كثافة الاحتمال المشتركة) تساوى :

$$g(x_1, x_2, ..., x_n; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2}$$

وبإضافة وطرح $\bar{x}$ نجد أن الطرف الأيمن يساوى

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\bar{x}_i-\bar{x})^2}\; e^{-\frac{n}{2\sigma^2}(\bar{x}-a)^2}$$

$$= k(x_1, x_2, \ldots / \bar{x})\, h(\bar{x}, a)$$

أى أن الوسط الحسابى ($\bar{x}$) هو مقدر كاف لتوقع التوزيع المعتاد .

١-٢-٤ التقدير بنقطة والتقدير بفترة (Point and Interval Estimation)

ذكرنا فيما سبق أن من أهم الأهداف التى يهتم بها الباحث ، تقدير معالم المجتمع كالوسط الحسابى والانحراف المعيارى من بيانات عينة عشوائية ، ويمكننا التمييز بين نوعين من التقدير :

- التقدير بنقطة .
- التقدير بفترة .

وسنقوم باستعراض هذين النوعين باختصار :

١ – التقدير بنقطة (Point Estimation)

يعد التقدير بنقطة النوع الأكثر شيوعًا من أنواع التقدير ، خاصة لدى غير الإحصائيين . والتقدير بنقطة هو تقدير لمعلمة المجتمع برقم واحد (أوقيمة وحيدة) ، مثلاً الوسط الحسابى للعينة ($\bar{x}$) هو تقدير بنقطة لوسط المجتمع (μ) . كذلك تقدير نسبة المجتمع من بيانات عينة (p) هو تقدير بنقطة لنسبة المجتمع (P) .

٢ – التقدير بفترة ثقة (Confidence Interval Estimation)

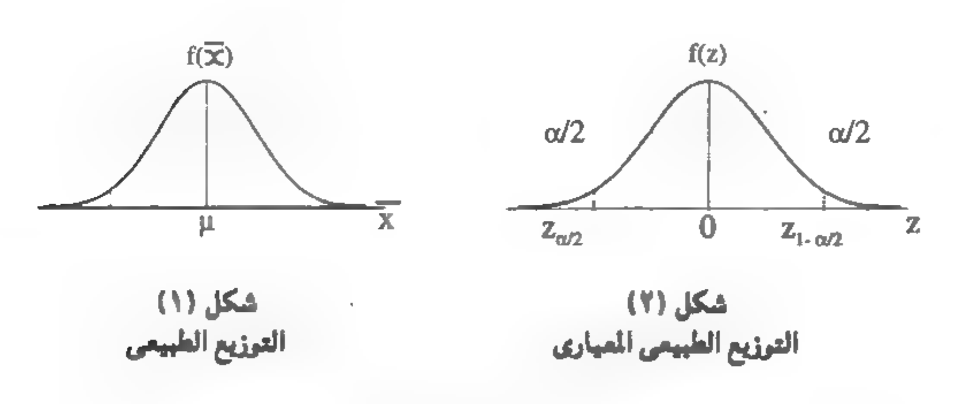
يسمى المدى الذى تقع فيه القيمة الحقيقية لمجتمع ما بدرجة ثقة معينة فترة الثقة ، والحد الأدنى والحد الأعلى لهذه الفترة حدود الثقة (Confidence Limits) . ونستطيع حساب الاحتمالات لفترة الثقة التى تحتوى على القيمة الحقيقية ، وتكون هذه الاحتمالات صحيحة فى حال استخدام المعاينة العشوائية البسيطة . كما أنه لا يمكن حساب حدود الثقة باحتمالات

صحيحة من بيانات عيّنات مسحوبة من مجتمعات مجهولة التوزيع . فإذا كان للتقدير توزيع طبيعي وكان الخطأ المعياري للتقدير معروفًا ، فإننا نستطيع معرفة احتمال وقوع خطأ في التقدير أكبر من أي قيمة أخرى ، لكن التقدير قد لا يتوزع بصورة طبيعية مما يجعل هذه الاحتمالات غير دقيقة . ولكن إذا كان حجم العينة كبيراً وكان التقدير غير متحيز ، فإننا نستطيع بمساعدة جداول التوزيع الطبيعي ومعرفة الخطأ المعياري للتقدير ، حساب فترة الثقة للقيمة الحقيقية لمعلمة المجتمع .

إذا كان (X) متغيراً عشوائيًا موزعًا طبيعيًا بمتوسط (μ) انحراف معياري (σ) فإن القيمة المعيارية

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad \text{............ (1 - 25)}$$

موزعة طبيعيًا بمتوسط (صفر) وانحراف معياري (١) . إن للوسط الحسابي للعينة العشوائية البسيطة ($\bar{x}$) المقدر من عينة حجمها (n) وحدة (من مجتمع له توزيع طبيعي وله متوسط μ وانحراف معياري σ) ، توزيعًا طبيعيًا متوسطه (μ) وتباينه (σ^2/n) ، لذا نجد أن للقيمة المعيارية توزيعًا طبيعيًا معياريًا .



شكل (١)
التوزيع الطبيعي

شكل (٢)
التوزيع الطبيعي المعياري

شكل رقم (٣)

منحنى التوزيع الطبيعي ومنحنى التوزيع الطبيعي المعياري

ويمكن القول كما يتضح من الشكل (٣) أن :

$$P\left(Z_{\alpha/2} \leqslant \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \leqslant Z_{1-\alpha/2}\right) = 1 - \alpha$$

حيث $Z_{\alpha/2}$ هى القيمة التى تسبقها مساحة $(\alpha/2)$ تحت المنحنى $Z_{1-\alpha/2}$ هى القيمة التى تسبقها مساحة $(1 - \alpha/2)$ تحت المنحنى و (α) هى المساحة المظللة تحت المنحنى خارج فترة الثقة ، و $(1-\alpha)$ هى درجة أو معامل الثقة ، ويمكن القول إن فترة الثقة للوسط الحسابى :

$$\overline{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad \text{........... (1 - 26)}$$

(قيمة $Z_{\alpha/2}$ سالبة وتستخرج من جدول التوزيع الطبيعى ، أما قيمته $Z_{1-\alpha/2}$ فهى موجبة مثلاً بمستوى ثقة (٩٥٪) نجد أن قيمة $Z_{1-\alpha/2}$ تساوى (1.96) $Z_{\alpha/2}$ تساوى (1.96-) كما هو موضح فى الملحق رقم (١) فى نهاية الكتاب .

إن تباين المجتمع $\sigma2$ غير معلوم فى كثير من الحالات ، لذا نستخدم توزيع ستيودنت (t Distribution) ، ويكون الخطأ المعيارى للمتوسط هو $\hat{\sigma}_{\overline{x}} = s \sqrt{n}$ وتكون القيمة المعيارية

$$t = \frac{\overline{x} - \mu}{\hat{\sigma}\ (\overline{x})}$$

موزعة حسب توزيع (t) بدرجات حرية (n-1) وتكون فترة الثقة فى حالة السحب مع الإعادة :

$$\overline{x} + t_{(\alpha/2, n-1)} \frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{(1-\alpha/2,\, n-1)} \frac{s}{\sqrt{n}} \qquad \text{........... (1 - 27)}$$

حيث :

$$s^2 = \frac{1}{n-1} \sum_{i-1}^{n} (x_i - \overline{x})^2$$

وتؤخذ قيم (t) من جدول توزيع ستيودنت (t) بمستوى ثقة معين ٪ $(1-\alpha)$ وبدرجات حرية (n-1) ، كما هو موضح فى الملحق رقم (٢) فى نهاية الكتاب .

أما فى حالة السحب مع عدم الإعادة ، تصبح فترة الثقة بعد إدخال معامل تصحيح المجتمع المحدود $(\frac{N-n}{N-1})$:

$$\bar{x} + t_{(\alpha/2,n-1)} \frac{s}{\sqrt{n}} \sqrt{1-f} \leq \mu \leq \bar{x} + t_{(1-\alpha/2,n-1)} \frac{s}{\sqrt{n}} \sqrt{1-f} \qquad \text{........ (1 - 28)}$$

(يلاحظ أننا استخدمنا $\frac{s}{\sqrt{n}} \sqrt{1-f}$ حيث تم الحصول على هذا المقدار من $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ بعد ضربها فى معامل تصحيح المجتمع المحدود :

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} \frac{N-n}{N-1} = \frac{S^2}{n} (\frac{N-n}{N})$$

لأنّ

$$\sigma^2 = \frac{N-1}{N} S^2$$

(s^2) هو مقدرغير متحيز لـ S^2 عندما يكون تباين المجتمع مجهولاً ، لذا وضعنا الانحراف المعيارى للعينة (s) فى صيغة فترة الثقة .

وهكذا نلاحظ أنه لاستخراج فترة الثقة لابد من تقدير الخطأ المعيارى ($\hat{\sigma}_{\bar{x}}$) أو تباين العينة (s^2) فى حالة عدم معرفة تباين المجتمع (σ^2) أو التباين المعدل للمجتمع (S^2) .

ويمكننا القول لتوضيح مفهوم حدود الثقة ، لو سحبنا عدداً كبيراً من العينات ذات الحجم (n) مفردة من المجتمع نفسه ، وحسبنا حدود الثقة لكل عينة ، فإن ٩٥٪ (إذا كانت ($\alpha = 0.05$)) من هذه الحدود لا بد أن تحتوى على متوسط المجتمع (μ) .

١-٤ أساليب جمع البيانات .

تتطلب مرحلة جمع البيانات ، تحديد الأسلوب المناسب لجمع البيانات ، لذا لابد لنا من التعرف على أساليب جمع البيانات (التى تسمى أساليب الحصر) ، وذلك بهدف التركيز على أسلوب المعاينة موضوع هذا الكتاب .

يعد تحديد الأسلوب المناسب لجمع البيانات من أصعب المشكلات التى يواجهها مصمم البحث . ويتوقف اختيار الأسلوب المناسب على عدد من المعايير :

- الدقة المطلوبة إذ يمكن استخدام أسلوب الحصر الشامل عندما تتوافر جميع الإمكانات المطلوبة والوقت الكافى ، ونريد الحصول على بيانات دقيقة وشاملة (مثلاً ، التأكد من جودة مظلات الجنود وسلامتها) .
- طبيعة الظاهرة التى نعالجها ومدى تجانس الوحدات الإحصائية ، إذ يفضل استخدام أسلوب المعاينة عندما يوجد تجانس بين هذه الوحدات ، خاصة إذا كان حجم المجتمع كبيراً أو يمكن تقسيمها فى مجموعات متجانسة .
- الإمكانات المادية والبشرية المتوافرة ، إذ لا يمكن استخدام أسلوب الحصر الشامل عندما لا تتوفر هذه الإمكانات .
- الوقت المخصص للبحث إذ يفضل استخدام أسلوب المعاينة عندما نريد الحصول على النتائج بسرعة .

إن اختيار أسلوب جمع البيانات المناسب ، يتوقف على المعايير السابقة ، لذا يجب اختيار الأسلوب المناسب الذى يعطى أكبر دقة ممكنة فى الوقت المحدد وذلك باستخدام الإمكانات المادية والبشرية المتوافرة .

ويمكننا التمييز بين ثلاثة أساليب لجمع البيانات :

- أسلوب الحصر الشامل .
- أسلوب الحصر الجزئى أو شبه الحصر .
- أسلوب المعاينة .

وسنقوم فى هذا الفصل بشرح مختصر لهذه الأساليب موضحين تعريف ومزايا وعيوب كل منها . وسنقوم فى الفصول القادمة بالتوسع فى دراسة أسلوب المعاينة .

١ - ٤ - ١ أسلوب الحصر الشامل (Complete Census or Complete Enumeration)

يعرف أسلوب الحصر الشامل بأنه أسلوب جمع البيانات الذى ندرس فيه حالة جميع وحدات المجتمع موضوع الدراسة دون استثناء . ويقضى هذا الأسلوب بجمع البيانات من جميع الوحدات الإحصائية دون استثناء أى منها . ويعد التعداد العام للسكان ، الذى ينفذ فى معظم الدول ، حصراً شاملاً لجميع السكان فى لحظة معينة ودولة معينة . كذلك يعد التعداد العام الزراعى حصراً شاملاً لجميع الحيازات الزراعية الموجودة فى دولة معينة . ويستخدم الحصر الشامل فى مجالات أخرى كالصناعة والتجارة ، وذلك لحصر المؤسسات الصناعية والتجارية حصراً شاملاً . وتهدف هذه التعدادات إلى الحصول على بيانات ومعلومات شاملة عن كل وحدة من وحدات المجتمع سواء كانت هذه الوحدة شخصاً أو أسرة أو مؤسسة أو أى وحدة أخرى .

ويستخدم أسلوب الحصر الشامل عندما نرغب فى الحصول على بيانات ومعلومات تفصيلية عن جميع الوحدات الإحصائية ، كذلك يستخدم هذا الأسلوب عندما يجهل الباحث طبيعة المجتمع بسبب عدم تنفيذ البحث فى فترة سابقة وعدم إمكانية اختيار عينة عشوائية تمثل المجتمع .

ويعد استخدام أسلوب الحصر الشامل ضرورياً فى بعض الحالات ، إذ تستخدم بياناته كأساس لتنفيذ بعض البحوث فى المستقبل لأنه يوفر الأطر اللازمة لاختيار وحدات العينة . ويتم استخراج معالم المجتمع والوصول إلى البيانات والمؤشرات الأخرى بالشكل والدقة المطلوبين من البيانات التى يتم جمعها بهذا الأسلوب . ويتصف أسلوب الحصر الشامل بعدد من المزايا والعيوب ، أهمها :

أ - مزايا أسلوب الحصر الشامل :

- الحصول على بيانات عن جميع الوحدات الإحصائية ، ويساعد ذلك على دراسة الظاهرة بشكل شامل . مثلاً نستطيع دراسة خصائص السكان الذين يقطنون فى بلد ما وتوزيعاتهم حسب السن والجنس والجنسية والحالة الزواجية والتعليمية وغيرها على مستوى الفرد والأسرة والدولة ككل .
- استخراج أهم معالم المجتمع ، مثلاً نستطيع باستخدام أسلوب الحصر الشامل للسكان حساب متوسط العمر والتباين وغيرها من معالم المجتمع التى تستخدم لأغراض التحليل الإحصائى .
- يساعد على إعداد إطار شامل لجميع وحدات المجتمع (قائمة بأسماء وعناوين الوحدات الإحصائية وأهم المعلومات الأخرى المتعلقة بها) ، وذلك لاستخدامه فى البحوث التى تنفذ

باستخدام أسلوب المعاينة . مثلاً من بيانات التعداد العام للسكان والمساكن ، يمكننا تكوين إطار الأسر وإطار المساكن التى تستخدم لتنفيذ البحوث التى تنفذ باستخدام أسلوب المعاينة كبحوث تكاليف المعيشة والعينة السكانية وغيرها .

- إمكانية استخدام الحصر الشامل فى حالة عدم توافر معلومات مسبقة عن الظاهرة المدروسة .

وتشجع هذه المزايا الباحثين على استخدام أسلوب الحصر الشامل خاصة فى فترات زمنية متباعدة ، وذلك لتكوين الأطر وحساب أهم معالم المجتمع ، خاصة إذا لم تتوافر بيانات مسبقة عن الظاهرة التى ندرسها .

ب - عيوب أسلوب الحصر الشامل :

يتصف أسلوب الحصر الشامل بعدد من العيوب التى تحدّ من استخدامه باستمرار . لذا يكتفَى بتنفيذه فى فترات متباعدة (كل خمس أو عشر سنوات) خاصة إذا كان حجم المجتمع كبيراً ، وأهم هذه العيوب هى :

- عدم إمكانية استخدامه إذا كان حجم المجتمع كبيراً أو لا متناهيًا ، مثلاً من الصعب إجراء حصر شامل للأشجار الموجودة فى إحدى الغابات ، أو رؤوس الأغنام الموجودة فى البادية أو غيرها .
- يتطلب هذا الأسلوب إمكانيات مالية وبشرية وفنية ضخمة لا يمكن توفيرها باستمرار . لذا يفضل استخدامه فى فترات زمنية متباعدة .
- يتطلب هذا الأسلوب وقتًا طويلاً فى جميع مراحله (التصميم وجمع البيانات والتبويب والوصف والتحليل والطباعة) .
- استحالة استخدامه فى بعض الحالات التى تؤدى إلى تلف الوحدات الإحصائية كفحص دم المريض بأجمعه ، لأنه يؤدى إلى وفاة المريض . وفحص جميع علب الحليب لأنه يؤدى إلى تلفها .
- الوقوع فى بعض الأخطاء نتيجة للتصميم الخاطئ للبحث ، أو للإرهاق الذى يصيب جامعى البيانات بسبب ضخامة عدد الوحدات المطلوب حصرها حصراً شاملاً .

يلاحظ مما سبق ، عدم إمكانية استخدام أسلوب الحصر الشامل فى بعض الحالات ، وعندما يستخدم هذا الأسلوب نجد أنّ تنفيذه يتم فى فترات زمنية متباعدة .

ويستخدم أسلوب الحصر الشامل فى مجالات متعددة ، وأهم البحوث التى تنفذ بأسلوب الحصر الشامل هى :

- التعداد العام للسكان والمساكن للحصول على البيانات المتعلقة بخصائص السكان وتوزيعاتهم المختلفة والبيانات المتعلقة بالمساكن .
- التعداد العام الزراعى للحصول على بيانات متكاملة عن النشاط الزراعى كالإنتاج والآلات والأراضى الزراعية وغيرها .
- التعداد العام الصناعى للحصول على بيانات شاملة متعلقة بالمصانع وغيرها من المؤسسات الصناعية .
- البحوث الميدانية التى تستخدم لأغراض البحث العلمى ، خاصة إذا كان حجم المجتمع صغيراً .
- المجالات الخطيرة التى تتطلب إجراء حصر شامل لجميع وحدات المجتمع .

١-٤-٢ أسلوب الحصر الجزئى (أو شبه الحصر) (Semi Enumration)

يستخدم أسلوب الحصر الجزئى (الذى يسمى أيضًا أسلوب شبه الحصر أو أسلوب البتر) فى مجالات متعددة، خاصة لحصر المؤسسات والمصانع الصغيرة والعاملين فى الصناعات الحرفية التى يكون عدد وحداتها كبيرًا ومساهمتها بالإنتاج قليلة إذا قورنت بمساهمة المصانع أو المؤسسات الضخمة .

عندما تتركز الظاهرة موضوع الدراسة ، فى عدد قليل (نسبيًا) من الوحدات الإحصائية ، نقوم بحصر هذه الوحدات حصراً شاملاً وتسمى هذه الوحدات «الوحدات المحصورة» . أما باقى الوحدات فإنها قليلة الأهمية لصغر مساهمتها على الرغم من ضخامة عددها إذا قورنت بالوحدات المحصورة ، لذلك نستغنى عن إدخالها فى البحث ونقوم بتقدير مساهمة هذه الوحدات باستخدام إحدى طرق التقدير المناسبة . وتسمى هذه الوحدات "الوحدات المبتورة" .

ويتطلب أسلوب الحصر الجزئى ، وجود دراسة نموذجية تم تنفيذها فى الفترة السابقة لمعرفة وتحديد الوحدات التى تتركز فيها الظاهرة ، وذلك للوصول إلى قاعدة عامة للتمييز بين الوحدات المحصورة والوحدات المبتورة ، وتحديد نسبة مساهمة كلا النوعين فى قيمة المتغير . ولتوضيح هذا الأسلوب ، نورد التطبيق الآتى :

تطبيق (١-٦)

أجريت فى عام ١٩٩٦م دراسة لتقدير إنتاج ودخل المؤسسات الصناعية والمنتجين الآخرين الذين يعملون فى صناعة الأحذية . وقد تبين أنّ (٣٪) من إجمالى عدد المؤسسات والمنتجين البالغ عددهم (١٠,٠٠٠) وحدة تساهم بحوالى (٨٥٪) من إجمالى إنتاج هذه المصانع . وقد بلغ إجمالى الإنتاج فى قطاع صناعة الأحذية (٤٥٠) مليون زوجٍ من الأحذية . وفى عام ١٩٩٧م ، تقرر اتباع أسلوب شبه الحصر لتقدير الإنتاج فى هذا القطاع فتم حصر المؤسسات الكبيرة البالغ عددها (٣٠٠) مؤسسة حصراً شاملاً وبلغ إنتاجها (٤٠٠) مليون زوجٍ . ما هو تقدير إجمالى إنتاج قطاع الأحذية فى عام ١٩٩٧م ؟

الحـــل

- إن نسبة تمركز الوحدات المحصورة حصراً شاملاً فى عام ١٩٩٦م :

$p = 300 / 10000 = 0.03$ i.e 3%

- نسبة الوحدات المبتورة فى عام ١٩٩٦ :

$q = 1 - p$
$= 1 - 0.03 = 0.97$ i.e 97%

- مساهمة الوحدات المحصورة فى إنتاج عام ١٩٩٦م :

$Y_{1996} = 450 \times 0.85 = 382.5$ Millions

ومساهمة الوحدات المبتورة فى الإنتاج لعام ١٩٩٦م :

$X_{1996} = 450 - 382.5 = 67.5$

- بافتراض ثبات النسب السابقة المتعلقة بتمركز الوحدات المحصوره والمبتوره ، نقدر مساهمة المؤسسات المبتوره باستخدام عدة طرق أبسطها الطريقة الآتية وذلك فى عام ١٩٩٧م :

400 مليون تقابل نسبة 85%
X_{1997} تقابل نسبة 15%

$$X_{1996} = 400 \times \frac{0.15}{0.85} = 70.58$$

ويكون إجمالى إنتاج قطاع الأحذية :

$$Y = 400 + 70.58$$
$$= 470.58 \text{ Millions}$$

ويمكننا استخدام عدة طرق لتقدير مساهمة الوحدات المبتورة فى الإنتاج كطريقة معدلات النمو (الزيادة) وطرق التنبؤ ، وشرح هذه الطرق خارج عن نطاق كتابنا .

يلاحظ مما سبق أن من مزايا هذا الأسلوب ، توفير الوقت والجهد والنفقات المالية ، نظراً لاقتصار البحث على عدد قليل من الوحدات الإحصائية ، خاصة وأن الإطار الخاص بالوحدات المبتورة لا يمكن إعداده . ويعاب على هذا الأسلوب اعتماده على نسب ودراسات سابقة قد تتغير من فترة إلى أخرى . لذا يتم البحث عن عامل مهم يؤثر على المتغير كعدد المشتغلين مثلا لاستخدامه للحد من عيوب هذه الطريقة .

ويستخدم هذا الأسلوب كثيراً فى بحوث الصناعات الحرفية كصناعة السجاد والنسيج والأحذية والزجاج وغيرها ، وذلك لتقدير الإنتاج والقيمة المضافة وغيرها من المؤشرات الإحصائية الاقتصادية .

١-٤-٢ أسلوب المعاينة (Sampling)

يتضح مما سبق أن طبيعة المجتمع الإحصائى الذى نقوم بدراسته وطبيعة البيانات المطلوبة ، تفرض على الباحث إجراء البحث بأسلوب الحصر الشامل أو أسلوب شبه الحصر . كما أنه لاعتبارات مادية وفنية وبشرية ، يفضل الإحصائيون والباحثون تنفيذ الكثير من البحوث بأسلوب المعاينة ، حيث يتم اختيار عينة من الوحدات الإحصائية لتعميم نتائجها والوصول إلى خصائص المجتمع من نتائج العينة التى تم اختيارها باعتبارها ممثلة للمجتمع الذى اختيرت منه .

لقد شاع استخدام أسلوب المعاينة كأسلوب لجمع البيانات بسبب المزايا التى يتصف بها ، وهناك بعض العيوب التى تحد من استخدامه .

ونورد فيما يأتى أهم مزايا وعيوب أسلوب المعاينة :

أ - مزايا أسلوب المعاينة :

- يتطلب هذا الأسلوب إمكانات بشرية ومالية وفنية قليلة إذا قورنت بالإمكانات التى تتطلبها الأساليب الأخرى ·

- السرعة ، إذ يتطلب تنفيذ البحث واستخراج نتائجه وقتًا أقل من الوقت الذى تتطلبه الأساليب الأخرى كأسلوب الحصر الشامل ·

- إمكانية استخدامه فى الحالات التى لا يمكن فيها استخدام أسلوب الحصر الشامل ، خاصة تلك التى تؤدى إلى تلف الوحدات الإحصائية (المرونة) ·

- اختبار دقة أسلوب الحصر الشامل ، إذ يستخدم أسلوب المعاينة لاختبار دقة أسلوب الحصر الشامل ·

- الدقة ، إذ يرى الكثير من الإحصائيين ، أن أسلوب المعاينة يعطى نتائج أفضل وأدق من نتائج الأساليب الأخرى بسبب إمكانية تقليل الأخطاء نتيجة لتركيز الجهود على عدد قليل من الوحدات الإحصائية والتدريب على عملية جمع البيانات والخطوات الأخرى وإمكانية المتابعة بسهولة ·

- إمكانية الحصول على بيانات أكثر تفصيلاً ، إذ يمكننا زيادة عدد الأسئلة فى الاستمارة عند استخدام أسلوب المعاينة لصغر حجم العينة وإمكانية تخصيص وقت أطول لكل وحدة ، وهذا غير متاح فى أسلوب الحصر الشامل خاصة عندما يكون عدد وحدات المجتمع كبيراً ·

ب - أهم عيوب أسلوب المعاينة :

- عدم إعطاء بيانات ومعلومات شاملة عن جميع وحدات المجتمع ، إذ تقتصر على وحدات العينة ·

- عدم إمكانية استخدام أسلوب المعاينة فى حال عدم توافر الإطار المناسب لكثير من أنواع العينات ، خاصة أن توافر الإطار يعد من الأمور الضرورية عند اختيار العينات ·

- قد يؤدى أسلوب المعاينة إلى نتائج غير دقيقة فى بعض الحالات ، خاصة إذا كانت هناك أخطاء تتعلق بتصميم البحث ، أو تقدير معلمات المجتمع ·

وعلى الرغم من هذه العيوب التى تحد من استخدام أسلوب المعاينة ، نجد أن استخدامه أصبح أكثر شيوعاً إذا قورن بالأساليب الأخرى ·

١-٥ أنواع العينات ومجال استخدامها .

يمكننا تقسيم العينات إلى نوعين رئيسين :

١-٥-١ عينات احتمالية :

وهى تلك العينات التى يتم اختيار وحداتها بشكل عشوائى ، وتستخدم فيها نظرية الاحتمالات ، حيث يتم اختيار وحداتها بشكل متتالٍ وباحتمالات محددة . ويتم سحب وحدات العينة وفق طرق محددة تسمى طرق السحب العشوائى ، ولا تسمح للباحث بالتدخل شخصياً فى اختيار أية وحدة إحصائية . ويمكننا التمييز بين الأنواع الآتية للعينات الاحتمالية :

أ - العينة العشوائية البسيطة :

تعد العينة العشوائية البسيطة أبسط أنواع العينات ، لكنها أكثرها أصالة فى العشوائية . ويتم اختيار وحدات العينة على أساس إعطاء فرص متكافئة لجميع وحدات المجتمع فى الظهور ، وتستخدم فى المجتمعات ذوات البيانات المتجانسة .

ب - العينة الطبقية العشوائية :

عندما تكون بيانات المجتمع غير متجانسة ، ويوجد فروق بينها ، يتم تقسيم المجتمع إلى أقسام (طبقات) طبقًا لمعايير معينة ، بحيث تختلف هذه الطبقات فيما بينها فيما يتعلق بالخاصية التى ندرسها ، بينما نجد أن هناك تجانسًا بين قيم الخاصية فى الطبقة الواحدة . ويتم اختيار عدد من الوحدات عشوائيًا من كل طبقة ، وتشكل العينات الجزئية المختارة من الطبقات العينة الطبقية العشوائية . وتستخدم هذه العينة فى المجتمعات غير المتجانسة .

جـ - العينة المنتظمة :

يتم اختيار وحدات العينة المنتظمة على أساس تقسيم المجتمع إلى فترات عددها يساوى حجم العينة المطلوب اختيارها . ويحدد رقم الوحدة الأولى باستخدام إحدى طرق السحب العشوائى ، ويتم تحديد أرقام الوحدات الأخرى بإضافة طول الفترة إلى رقم الوحدة الأولى فيتحدد رقم الوحدة الثانية . ثم نضيف إلى رقم الوحدة الثانية طول الفترة فيتحدد رقم الوحدة الثالثة ، وهكذا نكرر العملية حتى نحصل على أرقام وحدات العينة المنتظمة .

وتستخدم العينة المنتظمة بشكل واسع نظراً لسهولة اختيارها ، خاصة فى مجالات اختبارات الجودة فى خطوط الإنتاج ، وفى المجتمعات التى لا تتعرض لتغيرات دورية ٠

د – العينة العنقودية :

يقسم المجتمع إلى وحدات أولية (تسمى عناقيد أولية) يتم اختيار عدد منها بشكل عشوائى ويتم حصر العناقيد المختارة حصراً شاملاً حيث نحصل على قيم العينة التى عددها يساوى عدد العناقيد المختارة وقيمة كل منها هى قيمة العنقود . ونستطيع أن نميز بين العينة العنقودية البسيطة (ذات المرحلة الواحدة) والعينة العنقودية ذات المرحلتين والعينة العنقودية ذات المراحل المتعددة ، حيث نقوم بحصر العناقيد المختارة فى المرحلة الأخيرة حصراً شاملاً .

ويستخدم هذا النوع من العينات إذا كان المجتمع كبيراً ، ولا يتوافر إطار شامل لجميع الوحدات ، وهناك اختلاف بين العناقيد الأولية ٠

وسندرس العينات الاحتمالية بالتفصيل فى الفصول القادمة ٠

١-٥-٢ عينات غير احتمالية :

هى العينات التى لا يتم اختيار وحداتها بشكل عشوائى ، وإنما يتم الاختيار وفقًا لمعايير معينه يتدخل فيها الباحث عند سحب الوحدات . وتقسم العينات غير الاحتمالية إلى نوعين رئيسين :

أ – العينة الحصصية (Quata Sample)

يتم اختيار وحدات العينة الحصصية من قبل الباحث حيث يستخدم توجيهات مصمم البحث وبعض المعلومات التى تساعده على اختيار الوحدات وعند الانتهاء من الاستمارات المخصصة له (حصته) ، تكون العينة التى تم اختيارها عينة حصصية ٠ ويلاحظ أن عملية الاختيار لم تتم بشكل عشوائى ، وإنما تعتمد على حكم العداد ومقدرته على اختيار الوحدات وفق تعليمات مسبقة معطاة له ٠

ويستخدم هذا النوع من العينات بشكل واسع فى بحوث استطلاعات الرأى العام التى يقوم بها معهد جالوب فى الولايات المتحدة الأمريكية قبل إجراء الانتخابات الأمريكية ٠

ب – العينة العمدية (القصدية) (Purposive Sampling)

يقوم الباحث فى العينة العمدية بإدخال بعض الوحدات بشكل متعمد لاعتقاده توافر صفات ومعايير معينة فى هذه الوحدات تؤثر على الخاصية المدروسة وذلك للتأكد من وقوعها ضمن وحدات العينة ، أى يتعمد الباحث إدخال بعض الوحدات ضمن العينة المختارة .

مثلاً عندما نرغب فى اختيار عينة من أصحاب المحلات ، للتأكد من سلامة إجراءات الدفاع المدنى ، ندخل بعض المحلات التى تبين فى الفترات السابقة عدم التزامها بالتعليمات المعطاة وذلك للتأكد من وقوع هذه المحلات ضمن وحدات العينة .

وهناك أنواع أخرى من العينات ، يندرج بعضها تحت أنواع العينات الاحتمالية كالعينة المساحية والعينة المزدوجة والأنواع الأخرى من العينات التى سيتم دراستها فى الفصول القادمة .

الفصل
الثانى

الخطوات الأساسية
لتصميم العينة وجمع البيانات

٢-١ خطوات تصميم العينة :

تسمى الخطوات التى تسبق عملية جمع البيانات ميدانيًا خطوات تصميم العينة أو خطوات المرحلة التحضيرية للبحث . وسنقوم بشرح أهم هذه الخطوات لأهميتها عند دراسة الموضوعات المتعلقة بالعينات .

٢-١-١ تحديد المشكلة :

إن الخطوة الأولى من خطوات تصميم العينة هى تحديد المشكلة التى نقوم بدراستها بهدف تعريفها تعريفًا واضحًا والتأكد من الفوائد المرجوة من البحث . وتنشأ المشكلة نتيجة تفاعل الإنسان مع البيئة التى يعيش فيها إذ كثيرًا ما تواجهه مشكلة عندما يكون أمام موقف غير واضح يحتاج إلى تفسير أو عندما يكون هناك نقص فى المعلومات أو الخبرة أو رغبة فى تقصى الحقائق أو الإجابة على أسئلة غامضة .

ويقصد بتحديد المشكلة صياغتها فى عبارات واضحة ومفهومة ومحددة ، تعبر عن مضمون المشكلة ومجالها ، وتفصلها عن كافة المجالات الأخرى* . وهناك طريقتان لصياغة المشكلة :

- صياغة المشكلة بعبارة لفظية تقريرية .
- صياغة المشكلة بسؤال واحد أو عدة أسئلة .

وتعد الخبرة العملية والقراءات والدراسات والأبحاث السابقة أهم مصادر الحصول على المشكلة . إذ تزود هذه المصادر الباحثين بمشكلات تستحق الدراسة .

ولتوضيح كيفية تحديد المشكلة ، نورد المثال التالى :

يرغب باحث فى تحديد العلاقة بين متغيرين هما السرعة وعدد حوادث المرور . يمكن صياغة المشكلة كما يلى :

- الصيغة اللفظية التقريرية :

«علاقة السرعة بعدد حوادث السيارات»

وإذا أردنا وضوحًا وتحديدًا أدق يمكن توضيح هذه العلاقة بالصيغة التالية :

«علاقة السرعة بعدد حوادث السيارات فى الطرق التى تربط المدن ببعضها»

- صياغة المشكلة بسؤال :

«ما أثر السرعة على عدد حوادث السيارات فى الطرق التى تربط المدن ببعضها؟ » .

---

* د. ذوقان عبيدات وآخرون : البحث العلمى ، ١٩٨٢م (ص ٦٨) .

٢-١-٢ أهداف البحث :

يعد تحديد الهدف الرئيسى للبحث وتحديد أهدافه التفصيلية ذا أهمية ، كبيرة وذلك لتحديد البيانات المطلوب جمعها واستخدامها من قبل الباحث لكسب ثقة المدلى بالبيانات ، ونجد فى المثال السابق أن الهدف العام للبحث هو الكشف عن العلاقة بين السرعة وحوادث المرور .

ويقوم الكثير من الباحثين بصياغة أهداف تفصيلية للبحث توضح بشكل تفصيلى الأغراض التى يرغب الباحث فى الوصول إليها ، وفى المثال السابق يمكن صياغة أهداف البحث كما يلى :

- التعرف على أسباب حوادث السيارات وعددها .
- تحديد نسبة حوادث السيارات بسبب السرعة .
- قياس أثر السرعة على عدد حوادث السيارات فى الطرق التى تربط المدن ببعضها .

ويمكن إضافة أية أهداف أخرى يجدها الباحث ضرورية ومساعدة للوصول إلى إجابات لأسئلة البحث .

٢-١-٣ عنوان البحث :

يجب صياغة عنوان البحث بشكل مختصر وبلغة سهلة يعبران عن المشكلة التى نقوم بدراستها . وفى المثال السابق يمكن وضع عنوان البحث بالصيغة التالية :

«أثر السرعة على عدد حوادث السيارات فى الطرق الخارجية» .

٢-١-٤ شمول البحث (حدود البحث) (Coverage)

يتطلب تنفيذ البحث بشكل جيد ، وضع الحدود التى يجب عدم تجاوزها ، وذلك بهدف تحقيق الأهداف المرجوة من البحث . إن شمول البحث (أى حدود البحث) تعنى تحديد المناطق الجغرافية والوحدات التى سيغطيها البحث وذلك وفق الخطة الموضوعة .

عند دراسة المشكلة السابقة ـ مثلاً ـ لا بد لنا من تحديد هل الدراسة ستغطى كافة الطرق الخارجية التى تربط بين جميع المدن أو ستقتصر على بعض الطرق ؟ وهل ستشمل جميع الحوادث التى تقع فى هذه الطرق أم تشمل فقط الحوادث الناجمة عن أسباب معينة ؟ ويمكن تحديد شمول البحث فى هذا المثال كما يلى :

يغطى البحث حوادث السيارات التى تقع فى الطرق الثلاثة التالية خلال شهر يناير (كانون الثانى) :

١ - طريق ....

٢ - طريق ....

٣ - طريق ....

ويلاحظ أن البحث سيغطى فى هذه الحالة جميع حوادث السيارات الصغيرة والكبيرة التى تقع خلال شهر يناير فى الطرق الثلاثة المحددة ، أما بقية الحوادث التى تقع فى الطرق الأخرى أو ضمن المدن فإنها لا تدخل فى البحث .

٢-١-٥ تعريف وحدة المعاينة والمجتمع الإحصائى :

لا بد من تعريف وحدة المعاينة والمجتمع الإحصائى الذى نقوم بدراسته تعريفًا واضحًا لا التباس به ، وذلك لجمع البيانات من الوحدات ذات العلاقة بدقة تامة .

لدراسة الإنتاج والمبيعات فى قطاع الصناعات النسيجية ـ مثلاً ـ ، نختار عينة من المنشآت تكون ممثلة للمنشآت التى تعمل فى هذا القطاع فى منطقة معينة .

نلاحظ فى هذا المثال أن المنشأة هى وحدة المعاينة ، وتعرف بأنها الشركة أو المؤسسة التى تقوم بتصنيع المنسوجات سواء كانت منشأة فردية أو تعود ملكيتها لعدد من الأشخاص أو المساهمين ...

ويكون المجتمع الإحصائى هو جميع المنشآت التى تعمل فى قطاع النسيج فى المنطقة التى نقوم بدراستها أى أن المجتمع يتكون من وحدات المعاينة جميعها .

إن تحديد وتعريف الوحدة الإحصائية والمجتمع بشكل واضح يساعد على جمع البيانات من الوحدات المحددة وإعداد الإطار على ضوء التعاريف المحددة .

٢-١-٦ إعداد الإطار الإحصائى :

إن عملية اختيار عينة من مجتمع ما ، تتطلب توافر قائمة بأسماء الوحدات الاحصائية وعناوينها وأهم المعلومات المتعلقة بها ، وتسمى هذه القائمة «الإطار» .

ويمكننا تعريف الإطار بأنه قائمة (أو سجل) تتضمن أسماء الوحدات الإحصائية للمجتمع (وحدات المعاينة) وعناوينها وأهم البيانات والمعلومات التى تتعلق بها ، ويكون الإطار على شكل قائمة أو بطاقات أو سجل أو خريطة أو غير ذلك .

ونورد فيما يلي أحد الإطارات كمثال :

إطار المؤسسات الصناعية فى مدينة ....

اسم المؤسسة	النشاط الرئيسى	العنوان	رقم الهاتف	عدد العمال
١ – شركة محمد سعيد	صناعة النسيج	المنطقة الصناعية (٥)	٤٣٨٦٥٤	٢٠
٢ – شركة على محمد	صناعة مواد غذائية	المنطقة الصناعية (٦)	٤٣٨٨٨٨	٤٠
٣ –	....	....	....	......
٤	....	....	....	......

هناك شروط يجب توافرها فى الإطار حتى يكون جيداً :

– شمول الإطار لجميع الوحدات الإحصائية دون استثناء (وحدات المجتمع) .

– عدم تداخل إطار لنشاط معين مع إطار نشاط آخر . عدم تداخل إطار المؤسسات التجارية – مثلاً – مع إطار المؤسسات الصناعية .

– عدم تكرار الوحدة الإحصائية فى الإطار الواحد .

– تحديث الإطار بإدخال التعديلات من حيث الإضافة أو الحذف أوالتعديل باستمرار .

– ترتيب الإطار بشكل يساعد فى الوصول إلى الوحدات بسهولة .

٢-١-٧ صياغة فروض البحث :

الفروض هى حلول مؤقتة أو تفسيرات مؤقتة يضعها الباحث لحل مشكلة البحث . ويمكن القول إن الفروض هى إجابات محتملة لأسئلة البحث ، وتوضع بشكل علاقة بين متغيرين أو أكثر .

ويمكن صياغة الفروض بإحدى طريقتين :

– الطريقة المباشرة وهى توضح وجود علاقة بين المتغيرين ، وتسمى الفروض فى هذه الحالة فروضاً مباشرة .

– طريقة الفرض الصفرى إذ تصاغ الفرضية بشكل ينفى وجود العلاقة ، وتسمى الفروض فى هذه الحالة فروضاً صفرية (Null hypothesis) .

ولابد للباحث من إثبات الفروض التى وضعها عن طريق اختبارها واتخاذ القرارات المناسبة ، كذلك يستطيع الباحث أن يثبت فروضه عن طريق الاستنباط أو الرؤية المباشرة .

ولتوضيح ما سبق نورد الأمثلة التالية :

- إذا كان سؤال البحث : ما أثر التدخين في الإصابة بسرطان الرئة ؟ . فتكون هناك عدة إجابات على هذا السؤال منها مثلاً :

يوجد علاقة قوية بين التدخين والإصابة بسرطان الرئة ٠و يمكن صياغة هذه الإجابة بشكل فرضية :

الطريق المباشرة : توجد فروق إحصائية (معنوية) بين المدخنين وغير المدخنين من حيث الإصابة بسرطان الرئة ٠

الفرض الصفرى : لا توجد فروق إحصائية (معنوية) بين المدخنين وغير المدخنين من حيث الإصابة بسرطان الرئة ٠

إن استخدام الفرض المباشر أو الفرض الصفرى يتوقف على مدى رغبة الباحث فى تأييد وجود الفرق ، فإذا كان أكثر ميلاً إلى وجود الفرق يضع الفرض المباشر ٠

- إذا سمعت صوتًا خارج المنزل ، فإن سؤال البحث فى هذه الحالة يكون : ما هو سبب الصوت ؟ هناك عدة إجابات ، مثل : حادث سيارة أو سرقة أو غيرها ٠ إن اختبار الفروض التى تمثل الإجابات المحتملة يكون عن طريق الرؤية المباشرة ، أى نشاهد ما يحدث خارج المنزل ، وقد يكون ذلك عن طريق الاستنتاج ٠

ويفضل عادة الفروض التى يمكن قياسها واختبارها ، أى التى تحتوى على متغيرات ٠

٢-١-٨ تحديد البيانات المطلوب جمعها :

يتم تحديد البيانات المطلوب جمعها على ضوء أهداف البحث وفروضه ، وطرق التحليل التى سيتم اتباعها ، وطبيعة الوحدات والمجتمع ٠ ويتم ذلك باستشارة مستخدم البيانات والباحث الذى يحللها ٠

٢-١-٩ تحديد نوع العينة المناسب وحجمها :

إن طبيعة الوحدات الإحصائية وحجمها ومدى تجانسها من حيث الظاهرة التى ندرسها والبيانات المطلوبة والنفقات المالية والإمكانات البشرية والفنية المتوافرة ، تساعد مصمم البحث على اختيار النوع المناسب من العينات كالعينة العشوائية البسيطة أو الطبقية أو المنتظمة أو العنقودية وغيرها ٠ ويجب اختيار نوع العينة المناسب بدقة لاختلاف النتائج من نوع لآخر بسبب اختلاف الأساليب المتبعة فى الاختيار والتقدير ٠

كما يتم تحديد حجم العينة المناسب حسب نوع العينة المستخدم ، وذلك باستخدام الصيغة الرياضية التي تختلف من عينة لأخرى وذلك بمستوى ثقة معينة بعد تحديد خطأ التقدير الذي نقبله وأحيانًا التكلفة .

وسيتم استعراض طرق تحديد حجم العينة في الفصول القادمة عند استعراض الموضوعات المتعلقة بكل نوع من أنواع العينات .

٢-١-١٠ تحديد طريقة جمع البيانات :

يمكننا التمييز بين أربع طرق لجمع البيانات :

- المقابلة (أو الاتصال المباشر) .
- المراسلة (أو البريد) .
- وسائل الاتصالات كالهاتف والتلكس والحاسوب والفاكس .
- الملاحظة .

وسنقوم باستعراض هذه الطرق باختصار نظراً لأهميتها .

١ - طريقة المقابلة (أو الاتصال المباشر) .

تعد المقابلة من أهم طرق جمع البيانات إذ تستخدم كثيراً في البحوث الميدانية خاصة تلك التي تنفذها الأجهزة الإحصائية ومراكز البحوث والباحثون المهتمون بجمع بيانات دقيقة مباشرة من المدلين بالبيانات .

ويتم جمع البيانات بإجراء المقابلة (الاتصال المباشر) بين الباحث والمدلي بالبيانات (المستجوب) حيث يقوم الباحث بطرح السؤال وتدوين الإجابة فور سماعها (أو بعد الانتهاء من المقابلة) .

خطوات إجراء المقابلة :

إن المقابلة فن قائم بذاته ، وهناك أساسيات يجب اتباعها عند إجراء المقابلة للوصول إلى البيانات والمعلومات المطلوبة بشكل جيد ، أي تحقيق نجاح المقابلة . ويمكننا تلخيص هذه الأساسيات بما يلي :

أ – الإعداد للمقابلة بشكل جيد عن طريق :

– تحديد أهداف المقابلة وطبيعة البيانات والمعلومات التى سيحصل عليها الباحث ، حتى يتمكن من إعداد الوسائل المناسبة للحصول على البيانات .

– تحديد الأفراد الذين سيقابلهم الباحث .

– تحديد الأسئلة والإجابات المحتملة وذلك بهدف الاستعداد لإجراء المقابلة .

– تحديد موعد المقابلة والتقيد بالموعد المحدد وأن يكون مناسبًا للمُستجوَب .

– الإعداد للمقابلة من حيث المظهر والملبس ووسائل النقل .

– تحديد مكان المقابلة .

– التدرب على إجراء المقابلة خاصة إذا كانت طبيعة البيانات ذات أهمية وكان الأشخاص الذين ستتم مقابلتهم ذوى مراكز حساسة ولا يمكن مقابلتهم بسهولة .

ب – تنفيذ المقابلة وفق الخطة المحددة :

– الوصول قبل موعد المقابلة بفترة لضمان عدم التأخر .

– اللباقة فى الدخول إلى المُستجوَب وفى التعامل مع الآخرين (السكرتير ، المُستجوَب ، ...) .

– البدء بحديث ودى ثم توضيح أهداف المقابلة وطرح الأسئلة وإعطاء الوقت الكافى للإجابة وعدم إحراج المُستجوَب .

– تدون الإجابات بخط واضح وألا يستغرق الباحث وقتًا طويلا فى تسجيل الإجابات ، ويمكن استخدام إشارات أو رموز للإجابات لتقصير الوقت (الاختزال) ويمكن أن تسجل الإجابات دون تعديل أو إضافات .

– الانصراف بلباقة مع تقديم الشكر على تعاون المُستجوَب .

مزايا وعيوب طريقة المقابلة :

المزايا :

– الحصول على بيانات دقيقة من المصادر المحددة .

– خلق الثقة بين الباحث والمستجوب وإقامة علاقات ودية بينهما ، تساعد على الحصول على إجابات دقيقة وضمان تعاون المستجوبين .

- توضيح الأسئلة للمستجوبين خاصة إذا كانت هذه الاسئلة تحتاج إلى شرح وتفسير .
- ضمان الحصول على إجابات جميع الاسئلة وجميع الاستمارات .

العيوب :

- تتطلب المقابلة نفقات مالية وإمكانات بشرية ضخمة قد لا تتوافر في كثير من الأحيان خاصة إذا كان عدد وحدات العينة كبيراً .
- تتطلب وقتاً طويلاً .
- تسبب في بعض الأحيان حرجاً للمستجوَبين خاصة إذا كانت الأسئلة تتطلب إجابات محددة كالأسئلة الشخصية مثلاً .

وعلى الرغم من عيوب طريقة المقابلة ، فإنها تستخدم كثيراً في الحياة العملية ، نظراً لمزاياها التي تجعل الكثير من الباحثين يفضلونها على الطرق الأخرى ، خاصة في الدول النامية بسبب تدني المستوى الثقافي والوعي الإحصائي لدى السكان .

٢ - طريقة المراسلة (أو البريد) .

يتم في هذه الطريقة إرسال الاستمارات (الاستبانات) إلى المستجوَبين بالبريد أو تسلم إليهم باليد حيث يقومون بقراءة الأسئلة والإجابة عنها بأنفسهم .

وتستخدم هذه الطريقة كثيراً في بعض البحوث التي تنفذها مصلحة الإحصاءات العامة التي تجمع بيانات عن الجهات الحكومية . كذلك تستخدم عندما يكون مستوى الوعي الإحصائي مرتفعاً كما هو الحال في الدول المتقدمة .

تتصف هذه الطريقة بالمزايا والعيوب التالية :

المزايا :

- توفير الوقت خاصة إذا كان عدد الاستمارات كبيراً .
- تتطلب هذه الطريقة إمكانات مالية وبشرية قليلة خاصة إذا قورنت بطريقة المقابلة .
- سهولة هذه الطريقة ولا تتطلب إجراءات متعددة كما هو الحال في طريقة المقابلة .
- الحصول على إجابات لا يمكن الحصول عليها بدقة بالطرق الأخرى (إجابات الأسئلة المحرجة) .

العيوب :

- إهمال الاستبانات المرسلة وقد يكون مصيرها سلة المهملات أو عدم وصولها إلى المستجوَب لعدم وضوح العنوان .
- تأخر وصول بعض الإجابات لذا تحتاج إلى متابعة مستمرة .
- عدم اكتمال إجابات بعض الاسئلة لعدم وضوحها أو الإحجام عن الإجابة عنها .

وعلى الرغم من هذه العيوب ، تستخدم هذه الطريقة بشكل واسع بسبب المزايا التى تتصف بها خاصة انخفاض التكلفة . وللحد من عيوب هذه الطريقة لا بد من أن ترفق الاستبانة بخطاب تفصيلى يوضح أهداف البحث وإرشادات الإجابة ، وأن تصاغ الأسئلة بوضوح .

٢ - استخدام وسائل الاتصالات (الهاتف ، التلكس ، الفاكس ، الحاسوب) :

تعد هذه الطريقة من أسرع طرق جمع البيانات إذ تستخدم للحصول على إجابات سريعة مثل : استطلاعات الرأى العام .

ويستخدم الهاتف كوسيلة لجمع البيانات إذ يُعد أسهل الطرق وأسرعها ، ولكن يعاب على هذه الطريقة لجمع البيانات التصنت (عدم السرية) أو تدوين البيانات بشكل خاطئ إذا كان الصوت غير واضح .

كذلك يستخدم البعض التلكس ، إذ يتم طرح السؤال أو الأسئلة ويتم الحصول على الجواب أو الأجوبة إما بشكل فورى أو بعد فترة من الزمن . وتمتاز هذه الطريقة بأن الإجابات مكتوبة ويمكن الحصول عليها بسرعة ، ويعاب عليها عدم توافر التلكس لدى معظم الوحدات الإحصائية .

أما الفاكس فإنه يعد من أفضل الطرق ، إذ ترسل صورة الاستمارة ويقوم بملئها المستجوَب وإعادتها بالفاكس وذلك بسرعة كبيرة .

ويستخدم الحاسب الآلى (الحاسوب) لجمع البيانات عن طريق شبكة الاتصالات ، ويتطلب ذلك اشتراك الباحث والمدلى بالبيانات بالحاسب وهذا غير متاح فى معظم الأحيان .

إن استخدام طريقة جمع البيانات باستخدام وسائل الاتصالات يعتمد على توافر هذه الأجهزة لدى الجهة المنفذة للبحث ولدى المستجوَبين ، لذا لا تستخدم فى البحوث التى تكون فيها وحدات المعاينة الأشخاص كالموظفين والأسر وغيرها .

٤ – طريقة الملاحظة (أو المشاهدة) :

يتم جمع البيانات وفق هذه الطريقة من قبل الباحث على ضوء ملاحظاته ومشاهداته لسلوك معين ، وذلك من خلال اتصاله مباشرة بالأشخاص أو الأشياء التي يدرسها ، أو من خلال اتصاله بالسجلات والتقارير التي أعدها الآخرون .

مثلاً ، يستطيع الباحث تدوين البيانات المتعلقة بالسكن على ضوء مشاهداته (فيلا أو شقة ...) دون الحاجة لسؤال صاحب السكن . وتستخدم هذه الطريقة عندما لا يحتاج السؤال إلى إجابة من المدلي بالبيانات لسبب ما ، كما هو الحال لدى مرضى الأمراض العقلية وغيرها . وتتصف هذه الطريقة بعدم إحراج المدلي بالبيانات ، والسهولة وإمكانية استخدامها في حالات معينة لا يستطيع فيها المدلي بالبيانات إعطاء بيانات دقيقة .

أما أهم عيوبها ، فهو عدم الدقة في بعض الأحيان نتيجة التخمين الخاطئ للباحث . كما أن بعض المفحوصين يغيرون من سلوكهم عندما يشعرون بأنهم ملاحظون ، وقد يؤدي ذلك إلى نتائج خاطئة .

٢-١-١١ إعداد نماذج الجداول (الجداول الصماء) :

بعد تحديد الأسئلة والإجابات المحتملة ، يتم إعداد نماذج الجداول التي ستظهر فيها البيانات التي سوف تجمع . وتسمى هذه النماذج «الجداول الصماء» لأنها تحتوي على عناوين فقط ولا تتضمن أي أرقام . وتعد هذه الخطوة مهمة ، إذ توضح العرض الجدولي للبيانات التي سوف يتم جمعها وتبويبها ونشرها . وتعطى هذه الجداول أرقاماً متسلسلة لتسهيل الرجوع إليها . مثلاً ، إذا أردنا أن نجمع بيانات عن العمر والجنس للسكان ، يمكننا تكوين الجدول التالي :

جدول رقم ( )

توزيع سكان منطقة ... حسب العمر والجنس

العمر	الجنس		الإجمالي
	ذكور	إناث	
أقل من ٥			
٥ وأقل من ١٠			
١٠ " " ١٥			
١٥ " " ٢٠			
....			
٦٥ فأكثر			
المجموع			

٢-١-١٢ إعداد خرائط المسح والترقيم :

تعريف خريطة المسح .

تستخدم خرائط المسح فى البحوث الميدانية التى تنفذها الأجهزة الإحصائية وبعض المؤسسات لتسهيل الوصول إلى الوحدات الإحصائية لجمع البيانات منها .

وتعرف خريطة المسح بأنها الأداة التى تساعد الباحث على الوصول إلى الوحدات الإحصائية لجمع البيانات منها . وتتضمن الخريطة حدود الشوارع الرئيسية والشوارع الفرعية والأزقة (الحارات) والقطاعات الرئيسية والقطاعات الفرعية (البلوكات) وأرقام الوحدات . وتستخدم هذه الخرائط من قبل مصلحة الإحصاءات العامة والأجهزة الحكومية التى تنفذ بحوثًا كبيرة باستخدام أسلوب العينات ، خاصة إذا كانت طريقة جمع البيانات المستخدمة هى طريقة المقابلة .

أنواع خرائط المسح :

يتم إعداد عدة أنواع من خرائط المسح حسب فئات المستخدمين لها :

- خرائط المشرفين العامين وتتضمن أسماء الشوارع الرئيسية والفرعية وغيرها ، وتتضمن حدود المناطق التى تقع ضمن نطاق عمل المشرف .
- خرائط المفتشين وتتضمن حدود وتفاصيل المناطق التى تقع ضمن نطاق عمل المفتش .
- خرائط المراقبين وتتضمن حدود المناطق التى تقع ضمن نطاق عمل المراقب .
- خرائط العدادين وتتضمن أرقام الوحدات التى تقع ضمن نطاق عمل العدّاد وحدود منطقته .

وللخرائط أهمية كبيرة فهى تمنع الازدواجية بين عمل العدّادين ، وتساعد على الوصول إلى الوحدات المطلوبة، وهذا يؤدى إلى الحصول على بيانات دقيقة من الوحدات المحددة .

الترقيم :

يتطلب الوصول إلى الوحدات الإحصائية ترقيماً واضحاً للأسماء السكانية والمدن والقرى والأحياء والحارات والقطاعات والبلوكات والوحدات الإحصائية (كالمساكن) وذلك لتحديد موقع الوحدة الإحصائية على الخريطة .

ويقوم الجهاز الإحصائى بإعداد خرائط شاملة وتفصيلية عند تنفيذ التعداد العام للسكان والمساكن ، ويتم إدخال التعديلات بشكل مستمر للوصول إلى خرائط حديثة تستخدم فى البحوث الكبيرة التى تنفذ باستخدام أسلوب العينات وذلك حسب الوحدات المختارة .

وتتضمن الخرائط إشارات توضح كيفية المرور حول الأحياء وداخل الأزقة والقطاعات وأرقامها وحدودها لمساعدة الباحث على الوصول إلى الوحدات المطلوبة . وتوضع علامات فى بداية ونهاية الشوارع التى تحيط بالحى وتوضح أرقام الحى والقطاع والبلك وذلك على الشكل الآتى :

– العلامة التى توضع على بداية ونهاية حدود كل حى :

رقم الحى	رقم القطاع	رقم البلك

– العلامة التى توضع على بداية ونهاية حدود كل قطاع :

رقم القطاع	رقم البلك

– العلامة التى توضع للقطاع الفرعى (البلك) :

رقم البلك

ويتم الترقيم فى المدن والقرى كما يلى (ترقيم مقترح) :

– ترقيم المبانى على مستوى البلك بحيث يبدأ كل بلك برقم المبنى (١) .

– ترقيم المساكن على مستوى المبنى بحيث تبدأ أرقام تعداد مساكن المبنى برقم (١) .

– يكون تسلسل الأسر على مستوى القطاع بحيث يبدأ رقم أسر كل قطاع برقم (١) .

ويكون شكل العلامات التى توضع على مداخل المبانى والمساكن ..

– إذا كان المبنى يتكون من مسكن واحد :

رقم تعداد المسكن
____
رقم تعداد المبنى

– إذا كان المبنى يتكون من عدة مساكن (شقق) :

رقم تعداد أول مسكن – رقم تعداد آخر مسكن
____
رقم تعداد المبنى

إن الترقيم الواضح الجيد ، يساعد الباحث على الوصول إلى الوحدات المحددة له ، ويساعد على التقليل من الأخطاء التى تقع نتيجة لعدم جمع البيانات من الوحدات المختارة والمحددة على الخرائط التى توزع على الباحثين للاستدلال على أماكن الوحدات المختارة خاصــة إذا كان حجم العينة كبيراً وتقع فى أماكن متباعدة .

٢-١-١٢ تدريب المشتغلين :

يهدف التدريب إلى تنمية مهارات العاملين فى البحث على كافة مستوياتهـم وذلك باتباع التعليمات المحددة والعمل على أساس موحد .

أنواع التدريب :

يمكننا التمييز بين أنواع التدريب التالية :

أ - أنواع التدريب حسب مكان التدريب :

١ - تدريب مركزى يتم فى المركز الرئيسى للجهة المنفذة حيث يتم تدريب المشاركين فى البحث .

٢ - تدريب لا مركزى يتم فى مناطق متعددة .

ويحضر التدريب المركزى (إذا كان عدد العاملين فى البحث كبيراً) ، المشرفون والمفتشون العامون والمراقبون فى بعض الأحيان . ويتم إعداد التدريب اللامركزى للعدادين والمراقبين .

أما إذا كــان عدد المشاركين فى البحث قليلاً ، غالبًا مـا يتم تدريبهم مـركزيًا أى فى منطقة واحدة .

ب - أنواع التدريب حسب موضوعات التدريب :

ينقسم التدريب ، سواء كان مركزيًا أو غير مركزى ، إلى نوعين رئيسين :

١- التـدريب النظرى : ويشـمل هذا التـدريب مـحـاضـرات وحـالات عـمليـة على أهم الموضوعات المتعلقة بالبحث وهى :

- أهمية الإحصاء .
- أهمية البحث .
- التعليمات المتعلقة بجمع البيانات .

- تعريف الوحدة الإحصائية والمجتمع الإحصائى .
- محتويات الاستمارة والتعليمات المتعلقة بها .
- العلاقة بين المشتغلين .
- كيفية التعامل مع الجمهور .
- طريقة ترقيم الشوارع والمبانى .
- القواعد المالية والإدارية .
- كيفية استلام وتسليم الاستمارات .
- الأخطاء الممكن الوقوع فيها وكيفية تلافيها .

٢- التدريب العملى (الميدانى) : ويشمل هذا التدريب الموضوعات التالية :

- التدريب على ملء الاستمارة وكيفية مقابلة الجمهور .
- التعرف على المناطق التابعة للمشرف أو العداد .
- التدرب على ترقيم الشوارع والطرق وتحديد أماكنها على الطبيعة .

ويتم إعداد خطة التدريب على ضوء الاحتياجات من فئات المشتغلين وإمكاناتهم المهنية والثقافية . ويتم عادة إعداد دليل التدريب الذى يوضح خطة التدريب ، ويتضمن تفاصيل هذه الخطة المتعلقة بالموضوعات ومواعيد التدريب وأماكنها .

٢-١-١٤ الدعاية للبحث (الحملة الإعلامية)

أغراض الدعاية للبحث .

تهدف الحملة الإعلامية (للدعاية للبحث) إلى رفع مستوى الوعى الإحصائى لدى المدلين بالبيانات . ويمكننا تلخيص أهم أغراض الحملة الإعلامية بما يلى :

- تعريف الجمهور بأهمية الإحصاء وضرورة التعاون مع موظف الإحصاء .
- تعريف المدلين بالبيانات بأهمية البحث وأهدافه .
- إرشاد المدلين بالبيانات إلى طريقة الإدلاء بالبيانات أو ملء الاستبانة .
- توضيح سرية البيانات وطلب الدقة فى الإدلاء بالإجابات .

وسائل الدعاية للبحث .

تنقسم الوسائل التى تستخدم للدعاية للبحث إلى :

١ - وسائل مركزية كالإذاعة والتلفاز والصحف .

٢ - وسائل غير مركزية كالمنشورات التى توزع فى المناطق واللافتات والسيارات الإذاعية المتنقلة والهدايا والمحاضرات فى المدارس والجامعات وخطب أئمة المساجد ونشرة تعريفية ترفق بالاستبانة .

وتستخدم الوسيلة أو الوسائل المناسبة حسب طبيعة البحث والاستمارة المستخدمة والمدلين بالبيانات والإمكانات المالية المحددة للبحث . وقد تبين أن نجاح الحملة الإعلامية يؤدى فى كثير من الحالات إلى نجاح البحث والعكس بالعكس .

توقيت الحملة الإعلامية .

يتوقف توقيت الحملة الإعلامية على طبيعة البحث ، فهناك بحوث كبيرة تبدأ فيها الحملة الإعلامية قبل البحث بفترة كافية ، ويجب تكثيف مرات الحملة كلما اقترب موعد تنفيذ البحث .

٢-١-١٥ ميزانية البحث .

يتطلب تنفيذ البحوث الميدانية نفقات مالية تختلف من بحث لآخر حسب طبيعة البحوث وعدد وحدات المعاينة المختارة وتوزيعها الجغرافى . وتتطلب بعض البحوث مبالغ ضخمة يجب أخذ الموافقة عليها من قبل الجهات ذات العلاقة . لذا لا بد أثناء تصميم العينة ، من إعداد الميزانية التقديرية للبحث التى تتضمن جدولاً بتفاصيل أوجه النفقات ومبالغها المتوقعة ، والتى تشمل الرواتب والأجور ونفقات النقل والمسكن وطباعة الاستمارة والقرطاسية وغيرها (ويتم إعداد جدول تفصيلي لكل منها) والتاريخ المتوقع للإنفاق . ويتم إعداد جدول ميزانية البحث والذى يسمى أحياناً الخطة المالية للبحث وذلك بتقسيم النفقات إلى مجموعات متجانسة .

والجدول التالى ، يوضح ميزانية تقديرية لأحد البحوث :

الميزانية التقديرية لبحث ....

التاريخ المتوقع	البيان	المبالغ			
	مصاريف إدارية				
/١/١	رواتب وأجور	.			
/١/١ . .	قرطاسية	.			
/١/٢٠	طباعة استمارات	.		...	
	مصاريف نقل وسكن				
	أجرة سيارات	.			
/١/٣٠ ...	بنزين	. . . .			
/١/١ ...	فنادق وسكن	.			
-					
	تجهيز البيانات				
	حاسب آلى				
/٢/٥...	طباعة النتائـج	.			
/٢/٢٠ ...	. . .				
	المجموع			......	

٢-١-١٦ الخطة الزمنية للبحث .

يؤدى الوصول إلى النتـائج التى يهدف البحث إلى تحـقـيـقـهـا فى الوقت المحـدد إلى الاسـتـفـادة منهـا بشـكل أفضل . وللوصـول إلى النتـائج فى الوقت المحـدد ، لابد من إعداد الخطة الزمنية للبحث أو ما يسمى الجدول الزمنى للبحث .

ويتضمن الجدول الآتى تفاصيل الأعمال التى سيتم تنفيذها وعدد الأيام لإنجاز كل عمل وتاريخ البدء فى العمل وتاريخ الانتهاء منه .

إن إعداد هذا الجـدول يتطلب خـبـرة مـعـينة ، لأن أى خلل يقع فى أى خطوة يؤدى إلى الإخلال بالخطوات الأخرى ، ويؤدى ذلك إلى عدم التقيد بتاريخ البدء بجمع البيانات وتاريخ الوصول إلى النتائج .

الخطة الزمنية لبحث ....

البيان	عدد الأيام	البداية	النهاية	ملاحظات
أولا - تصميم العينة	٢٠	.../١/١	.../١/٢٠	
تحديد الأهداف	١	١/١	١/١	
إعداد الإطار	٥	١/١	١/٥	
			.	
			.	
ثانيا - جمع البيانات	١٠	٢/١	٢/١٠	
تدقيق الاستمارات	٥	٢/١٠	٢/١٥	
ثالثا - تبويب البيانات	١٠	٢/١٥	٢/٢٥	
رابعا - نشر النتائج				
خامسا - تحليل البيانات				

٢-٢ إعداد الاستمارة الإحصائية .

الاستمارة الإحصائية هي الأداة المستخدمة لجمع البيانات ، وهي عبارة عن وعاء كتابي يحتوى على الأسئلة التي تمكن الباحث من جمع البيانات التي يحتاجها ، وعلى فراغات لتدوين الإجابات والرموز . ويجب العناية بصياغة الأسئلة وتصميم الاستمارة بشكل يمكننا من الوصول إلى الإجابات بدقة متناهية .

٢-٢-١ أنواع الاستمارات .

يمكننا التمييز بين عدة أنواع للاستمارات حسب عدة معايير :

أ - أنواع الاستمارات حسب طريقة جمع البيانات :

١ - **الاستمارة** : هي التي يقوم فيها الباحث بطرح الأسئلة على المدلي بالبيانات ، ويقوم بتدوين الإجابات فيها فور تلقيها . ويتم جمع البيانات بواسطتها عن طريق المقابلة أو الملاحظة . وتتصف الاستمارة بعدة مزايا أهمها :

- الحصول على إجابات كاملة لجميع الأسئلة .
- توضيح الأسئلة غير المفهومة للمدلى بالبيانات من قبل الباحث .
- إمكانية استخدامها فى بعض الحالات التى تتطلب جمع البيانات بطريقة الملاحظة .
- توفير الوقت بالنسبة للمدلى بالبيانات .

أما أهم عيوبها فهى :

- ضخامة التكاليف المادية والإمكانات البشرية المطلوبة ، لأن هذا النوع يتطلب قيام الباحث بإجراء المقابلة بنفسه ، أو بالملاحظة وتدوين الإجابات .
- تسبب نوعًا من الإحراج خاصة إذا كانت الأسئلة محرجة وتتعلق بالأمور الشخصية .

وعلى الرغم من هذه العيوب ، يستخدم هذا النوع من الاستمارات كثيرًا ، خاصة فى الدول النامية بسبب انخفاض المستوى الثقافى والتعليمى ومستوى الوعى الإحصائى لدى السكان .

٢ - **الاستبانة :** وهى الأداة المستخدمة لجمع البيانات عن طريق المراسلة (أو الفاكس) .

وتختلف الاستبانه عن الاستمارة فى أن الاستبانة تحتوى على معلومات إضافية تساعد المدلى بالبيانات على ملء الأجوبة بوضوح ودقة متناهية . وتتلخص المعلومات الإضافية بما يلى :

- خطاب تغطية يوضح أهمية البحث وأهدافه وإرشادات ملء الاستبانة وكيفية إعادتها والتاريخ الأقصى للإعادة .
- التعاريف الرئيسية وشرح بعض الأسئلة إذا كانت تحتاج إلى ذلك .

تتصف الاستبانة بعدد من المزايا مما يشجع البعض على استخدامها ، وهى :

- تتطلب إمكانات بشرية ومادية قليلة لإرسالها بالبريد .
- الإسراع فى الوصول إلى الوحدات الإحصائية ، خاصة إذا كان عددها كبيرًا ومتناثرة فى مناطق متباعدة .
- عدم الإحراج إذ يستطيع المدلى بالبيانات الإجابة دون الشعور بالحرج وكتابة الإجابة بصراحة أكثر .

أما أهم عيوبها فهى :

- عدم وصول جميع الاستبانات وإهمالها من قبل المستجوب ، لذا تتطلب متابعة مستمرة .
- عدم اكتمال إجابات جميع الأسئلة إذ كثيرًا ما يلاحظ وجود أسئلة لم يجب عليها المستجوَب .
- تأخر وصول بعض الإجابات رغم المتابعة المستمرة ، وفقدان بعضها فى البريد .

ب – أنواع الاستمارات حسب درجة شمولها .

١ – الاستمارة الفردية : وهى الاستمارة التى تخصص لوحدة إحصائية واحدة . مثلاً لدراسة دخل وإنفاق عدد من الموظفين في إحدى الجهات ، تخصص استمارة لكل موظف ، وبذلك نحتاج إلى عدد من الاستمارات يساوى حجم العينة .

وتمتاز هذه الاستمارة بإمكانية الحصول على معلومات تفصيلية عن كل وحدة ، كما أنها تساعد على تحديث الإطارات وتكوينها خاصة إذا استخدم أسلوب الحصر الشامل .

أمــا أهم عيوبها فتتلخص فى ضخامة عدد الاستمارات ، خاصة إذا كان حجم العينة كبيراً ، وضخامة التكاليف المادية والجهود الكبيرة المتعلقة بتخزينها .

٢ – الاستمارة الجماعية : وهى الأداة المستخدمة لجمع البيانات لعدد (مجموعة) من الوحدات الإحصائية ، وذلك بتخصيص سطر لكل وحدة إحصائية . فى المثال السابق ، يمكننا تخصيص استمارة واحدة لكل إدارة من إدارات الجهة التى ندرسها ، ونخصص سطراً لكل موظف تدون عليه البيانات المطلوبة .

من مميزات الاستمارة الجماعية ، توفير الوقت والجهد والنفقات . أما أهم عيوبها فهو عدم القدرة على الحصول على معلومات تفصيلية عن الوحدة الإحصائية .

٢-٢-٢ الأسس الواجب مراعاتها فى الاستمارة الجيدة :

هناك عدد من الأسس الواجب مراعاتها عند تصميم الاستمارة :

أ – الأسس المتعلقة بشكل الاستمارة :

– أن يكون حجم الاستمارة مناسباً والورق المستخدم من الورق الجيد .

– ترك أماكن كافية للإجابات والترميز .

– ترتيب الأسئلة بشكل منطقي ويفضل تقسيمها إلى أقسام متجانسة والبدء بالأسئلة السهلة .

– أن يشمل الغلاف الجهة المنفذة للبحث وعنوان البحث والإشارة إلى سرية المعلومات (فى حالة استخدام الاستبانة) .

– ترقيم الأسئلة بشكل يساعد على تبويب البيانات وسهولة الرجوع إلى أى سؤال .

ب - الأسس المتعلقة بمحتويات الاستمارة وصياغة الأسئلة :

- صياغة الأسئلة بعبارات واضحة وكلمات سهلة واستخدام الكلمات الشائعة المفهومة من المستجوبين .
- الاقتصار على الأسئلة الهامة والمنطقية .
- صياغة الأسئلة القصيرة المرتبطة بالمعنى .
- صياغة الأسئلة بحيث تكون إجاباتها محددة وبسيطة ولا تتطلب عمليات حسابية مطولة أو تستدعى ذاكرة قوية .
- تحديد وحدات القياس (كغ ، طن ، دولار ، ...) عند الحاجة .
- تجنب الأسئلة المحرجة والأسئلة التى تستدعى الكذب .
- تجنب الأسئلة المفتوحة قدر الإمكان خاصة فى البحوث التى يتكرر تنفيذها .
- وضع بعض الأسئلة التى توضح صدق المجيب .

إن اتباع الأسس السابقة يساعد على تصميم استمارة جيدة تؤدى إلى نجاح عملية جمع البيانات . ويوضح الملحق رقم (٣) نموذجاً لإحدى الاستمارات .

٢-٢-٣ ترميز الاستمارة :

تعريف الترميز ودليل الترميز :

تتطلب عملية إدخال البيانات على الحاسوب ، التعبير عن الإجابات برموز معينة لاستخدامها لأغراض التبويب أو التحليل .

والترميز هو «التعبير عن الإجابة برمز ما قد يكون رقماً أو حرفاً أو لفظاً» ، ويتم إعداد ما يسمى دليل الترميز للمساعدة على استخدام الرمز المناسب وتوحيد الرموز لجميع المبوبين .

يعرف دليل الترميز بأنه عبارة عن «قائمة تتضمن تفاصيل الإجابات (المحددة أو المحتملة) والرموز المقابلة لها» وذلك للاستعانة بها عند إدخال البيانات والمعلومات إلى الحاسوب .

أنواع الترميز :

يمكننا التمييز بين أنواع الترميز التالية ، لاستخدام الترميز المناسب الذى ينسجم مع إجابات الأسئلة :

أ – الترميز الرقمى : نرمز للإجابة برقم من الأرقام مثلاً نستخدم (١) للتعبير عن الإجابة «نعم» و (٢) للإجابة «لا» .

	١	نعم		٢	لا

ب – الترميز الحرفى : نرمز للإجابة بحرف من الحروف الهجائية مثلاً نستخدم (ذ) إذا كان المجيب ذكراً و (أ) للأنثى .

	ذ	ذكر		أ	أنثى

ج – الترميز اللفظى : نرمز للإجابة بلفظ معين للتعبير عن إجابة مؤلفة من عدد من الكلمات .

مثلاً نستخدم كلمة «أوافق» إذا كانت الإجابة أوافق على زيادة عدد الموظفين و «غير موافق» إذا كانت الإجابة عكس ذلك .

☐ أوافق ☐ غير موافق

ويقوم المجيب باختيار إحدى الإجابتين بوضع إشارة (√) أمام إحداهما .

د – الترميز الرقمى الحرفى : نستخدم رقماً وحرفاً للإشارة إلى إجابة مزدوجة ، مثلاً نستخدم الرقم كرمز للمدينة والحرف للجنس .

تكون الرموز إذا كان لدينا مدينتان :

١ ذ إذا كان المجيب ذكراً من المدينة الأولى .

٢ أ إذا كان المجيب أنثى من المدينة الثانية .

وهكذا .

إن الترميز يساعد كثيراً على تسهيل عملية إدخال البيانات والتعامل بها ، ويجب استشارة أحد المتخصصين فى الحاسوب عند ترميز الاستمارة وتصميمها .

طرق الترميز :

يمكننا استخدام طريقة أو أكثر من طرق الترميز التالية :

أ - نسخ الرموز وطباعتها بجانب الإجابات المحتملة ، أى تطبع الرموز على الاستمارة :

١		نعم	٢		لا

ب - تخصيص أماكن للرموز وترمز الإجابة على الاستمارة بعد جمع البيانات :

		نعم			لا

جـ - ترميز الإجابات فى كشوف خاصة تحتوى على الإجابات فقط مثلاً :

رقم الاستمارة	الجنس	المدينة
١	ذ	١
٢	ذ	٢
٣	أ	١
.	.	.

د - ترميــز البيانــات فى بطاقات ترميز حيث تخصص بطاقة لكل حــالة (ســؤال) على حدة ، مثلاً :

رموز توزيع المجيبين حسب الجنس

رقم الاستمارة	الجنس
١	ذ
٢	أ
٣	ذ
.	.

وتستخدم كل طريقة حسب نوع الأسئلة ، إذ نجد أن الطريقة (أ) تستخدم إذا كانت الأسئلة مغلقة (أى تحدد إجاباتها بشكل مسبق) حيث يتم طباعة الرمز بجانب الإجابة وهى الأكثر استخداما . أما إذا كانت الأسئلة مفتوحة (يترك فراغات لإجابات المدلين بالبيانات) فيتم استخدام الطريقة (ب) حيث يتم الترميز بعد وصول الإجابات . وتستخدم الطريقتان (جـ) ، (د) لتسهيل عملية تبويب البيانات .

٢-٣ البحث التجريبي (العينة الاستطلاعية)

تسمى العينة الاستطلاعية في بعض البحوث العينة القبلية أو البحث التجريبي · وكما يستدل من هذه التسميات ، نجد أن الهدف الأساسي للعينة الاستطلاعية هو اختبار وسيلة جمع البيانات أي الاستمارة ، وجميع الخطوات المتعلقة بتصميم البحث · ويتم اختيار عينة من الوحدات توزع عليها الاستمارات لملئها، أو يقوم الباحثون بملئها ووضع الملاحظات عليها لتعديل الاستمارة إذا كانت تحتاج إلى تعديل ، أو إجراء التعديلات اللازمة على خطوات تصميم العينة إذا كانت تحتاج إلى ذلك ، وبعد الانتهاء من إدخال التعديلات يتم طباعة الاستمارة بشكلها النهائي ·

ويتم التركيز في العينة الاستطلاعية على اختبار محتويات الاستمارة ، والتأكد من سهولة الأسئلة ووضوحها ، وفيما إذا كانت الأماكن المخصصة للإجابات كافية ، ... كذلك يتم التأكد من مدى نجاح الحملة الإعلامية للبحث ، ومن فعالية تدريب المشتغلين · كما يستفاد من هذه العينة في توفير بيانات أولية تستخدم لتقدير حجم العينة أو لأغراض أخرى .

وكثيراً ما يقوم الباحثون بقياس مدى صدق الاستمارة باستخدام عدة طرق ، منها ـ مثلاً ـ توزيع الاستمارة على المجموعة نفسها مرتين ، وملؤها ومن ثم حساب معامل الارتباط بين الأجوبة الأولى والثانية ·

لقد ثبت عمليًا أهمية العينة الاستطلاعية ، حيث تم تعديل الكثير من خطوات تنفيذ البحث على ضوء الملاحظات التي تم استنتاجها قبل تنفيذ البحث ·

٢-٤ جمع البيانات وتدقيقها .

٢-٤-١ جمع البيانات حسب الخطة المحددة .

يبدأ الباحثون (العدادون) بعملية جمع البيانات في الموعد المحدد ، وتستمر لفترة زمنية كما هو محدد في الخطة الزمنية ، وذلك إذا كانت طريقة جمع البيانات المستخدمة هي المقابلة أو الملاحظة ·

أما إذا كانت الطريقة المستخدمة لجمع البيانات المراسلة ، فيتم إرسال الاستبانات بالبريد أو تسلم باليد من قبل المراسلين ·

وعند الانتهاء من عملية جمع البيانات ، ترسل الاستمارات إلى الإدارة المركزية · ويتم عادة تدقيق الاستمارات أثناء عملية جمع البيانات وبعد الانتهاء منها .

٢-٤-٢ تدقيق الاستمارات (المراجعة) .

للتأكد من دقة البيانات التي سنحصل عليها ، لابد من مراجعة البيانات التي جمعت في الاستمــارة للتأكد من صحتها قبل تبويبها . وتسمى هذه العملية تدقيق الاستمارات أو المراجعة . ويمكننا التمييز بين نوعين من المراجعة :

١ - **المراجعة الميدانية** : وهي المراجعة التي تتم أثناء عملية جمع البيانات وتتم من قبل العداد والمراقب والمفتش .

يقوم العداد بعد طرح الأسئلة على المدلي بالبيانات بمراجعة سريعة للإجابات التي دونها للتأكد من وضوحها وعدم نسيان أية إجابة . ويقوم المراقب بمراجعة الاستمارات التي تسلم إليه يوميًا من العداد ، ويقوم بزيارة بعض المستجوَبين للتأكد من قيام العداد بجمع البيانات منهم ، كما يقوم المفتش أيضًا باختيار عدد من الاستمارات ومراجعتها قبل إرسالها للإدارة المركزية .

٢ - **المراجعة المكتبية** : وهي المراجعة التي تتم من قبل المراجعين بعد وصول الاستمارات أو الاستبانات إلى الإدارة المركزية ، ويتم أحيانًا ترميز بعض الإجابات أثناء المراجعة . ويفضل تقسيم الاستمارة إلى عدة أقسام يقوم المراجع بتدقيق أحد الأقسام في جميع الاستمارات لتسهيل عملية المراجعة .

أهداف عملية تدقيق الاستمارات .

- التأكد من صحة ودقة البيانات التي جمعت .
- التأكد من عدم تعارض الإجابات والبيانات ومن تماسكها .
- التأكد من تسجيل الإجابات وفق التعليمات المعطاة وأن جميع الأسئلة قد أجيب عنها .
- التأكد من تسجيل البيانات بشكل يساعد على الترميز وإدخال البيانات على الحاسوب .
- التعرف على ملاحظات العدادين والمشرفين عليهم للاستفادة منها .
- قبول الاستمارة إذا كانت مستوفية للتعليمات أو رفضها إذا لم تكن دقيقة ، والعمل على الرجوع إلى المدلي بالبيانات .

ولتدقيق الاستمــارات أهميــة كبيرة ، إذ كثيرًا ما تكتشف الأخطاء بشكل مبكر قبل تبويبها ، ويؤدي ذلك إلى نتائج دقيقة وشاملة لجميع الأسئلة الواردة في الاستمارة .

٢-٥ مصادر الأخطاء فى العينات وكيفية التقليل منها .

عند تنفيذ البحوث الإحصائية سواء كان أسلوب جمع البيانات المستخدم هو أسلوب الحصر الشامل أو أسلوب المعاينة ، هناك نوعان من الأخطاء التى قد نقع فيهما : أخطاء التحيز (Bias Errors) والتغيرات العرضية (Accidental variations) . إن التحيز هو عبارة عن انحراف متوسط جميع تقديرات معلمة المجتمع للعينات الممكنة عن قيمتها الحقيقية ، ويوجد صعوبة فى التخلص منه أو تقليل أهميته . أما التغيرات العرضية فيمكن ملاحظتها من الفروق التى تنتج إذا تكرر تنفيذ البحث ، وتنتج هذه التغيرات من اختلاف العدادين واختلاف تعاون المدلين بالبيانات والطقس وغيره من العوامل التى تؤدى إلى اختلاف النتائج إذا تكرر إجراؤه مرة أخرى .

إن الأخطاء التى قد نقع فيها عند استخدام أسلوب المعاينة كأسلوب لجمع البيانات تسمى أخطاء المعاينة الكلية ويمكن تقسيمها إلى نوعين من الأخطاء :

- خطأ المعاينة العشوائى .
- خطأ التحيز .

وسنقوم بدراسة هذين النوعين نظراً لأهميتهما وضرورة التقليل منهما عند استخدام أسلوب المعاينة .

٢-٥-١ خطأ المعاينة العشوائى (Random Sampling Error)

عند اختيار عينة عشوائية حجمها (n) وحدة من مجتمع حجمه (N) وحدة معاينه ، نجد أن هناك خطأ ينتج عن الاختلاف بين قيم الوحدات التى تتكون منها العينة وتلك التى لم تشأ الصدف أن ندخلها فى العينة وهذا الخطأ يسمى خطأ المعاينة العشوائى .

ويمكننا باستخدام الطريقة المناسبة لاختيار الوحدة ، تحديد متوسط أخطاء المعاينة العشوائية من نتائج العينة وتوزيعها . إن الحجم المتوسط لهذه الأخطاء يعتمد على حجم العينة ، ومدى تشتت مفرداتها ، والإجراءات التى استخدمت لاختيار الوحدات .

وإذا عالجنا موضوع الأخطاء بعيداً عن أخطاء التحيز ، فإن الطريقة الأسهل لزيادة دقة نتائج العينة ، هى زيادة حجمها وذلك للتقليل من خطأ المعاينة العشوائى ويمكن القول إن خطأ المعاينة العشوائى ، يتناسب عكسياً مع الجذر التربيعى لحجم العينة . لقد ذكرنا أن هذه الأخطاء تعتمد على تباين مفردات العينة ، ويتم التقليل من هذا الجزء من الأخطاء باتباع طريقة الاختيار المناسبة كالأسلوب الطبقى أو أسلوب المعاينة ذات المراحل المتعددة أو أسلوب استخدام المعلومات المتممة .*

* لمزيد من التفاصيل ، راجع
Yates F.: Sampling Methods for Census, and Surveys, Charles Grifin & co. Ltd, 1981 (p. 17).

ويمكننا تقدير خطأ المعاينة العشوائى ، إذا كنا نقدر أحد معالم المجتمع بحساب الانحراف المعيارى لمتوسطات العينات الممكنة الذى يسمى الخطأ المعيارى (Standard Error) ونستخدمه للحكم على دقة الوسط الحسابى فى المعاينات العشوائية وتقدير حجم العينة .

وسنعالج كيفية حساب الخطأ المعيارى لبعض أنواع المعاينات فى الفصول القادمة وذلك باستخدام تباين المجتمع أو تقديره من عينة إذ ليس ضرورياً معرفة جميع العينات الممكنة لاستخراج قيمة هذا الخطأ .

٢-٥-٢ أخطاء التحيز وأنواعها (Bias Errors)

عندما نستخدم أسلوب المعاينة لتقدير معلمات المجتمع ، فإن متوسط جميع التقديرات المحسوبة باستخدام مقدر معين للعينات الممكنة ، يجب أن يساوى قيمة المعلمة التى نقوم بتقديرها ، وفى حالة وجود فرق بينهما فإن هذا الفرق يسمى خطأ التحيز . ويعرف خطأ التحيز بأنه انحراف متوسط جميع تقديرات معلمة المجتمع للعينات الممكنة عن القيمة الحقيقية لهذه المعلمة . ويتصف التحيز بأنه ثابت القيمة وتوجد صعوبة فى التقليل أو التخلص منه . إن خطأ التحيز لا يقل إذا ازداد عدد وحدات العينة بينما نجد أن خطأ المعاينة العشوائية يقل إذا ازداد حجم العينة كما ذكرنا .

ويمكننا التمييز بين ثلاثة أنواع من أخطاء التحيز :

أ - خطأ التحيز فى الاختيار

ب - خطأ التحيز فى التقدير

جـ - خطأ التحيز الناتج عن التعريف الخاطئ لوحدة المعاينة .

وسنقوم باستعراض هذه الأنواع باختصار ، وبشرح كيفية التقليل منها .

أ - خطأ التحيز فى الاختيار .

يوجد عدة طرق لاختيار وحدات العينة ، تؤدى إلى ارتفاع خطأ التحيز .

- الاختيار غير العشوائى لوحدات العينة الذى يعتمد على مزاج الباحث دون اتباعه للتعليمات المعطاة له ، وعدم اتباع طرق الاختيار العشوائى التى سنتطرق إليها فى الفصول القادمة .
- تعتمد بعض طرق الاختيار على خاصية معينة قد تكون مرتبطة بالخاصية المدروسة ، كالاعتماد على دليل الهاتف لاختيار عينة من السكان لدراسة الدخل والإنفاق ، حيث نجد

أن من لديهم الهاتف هم من أصحاب الدخول الجيدة . لذا يؤدى استخدام هذه الطريقة من الاختيار إلى الوقوع فى خطأ التحيز .

- التحيز المقصود أو غير المقصود فى اختيار وحدات العينة ، إذ قد يقوم الباحث باختيار بعض الوحدات متعمداً إدخالها ، أو غير متعمد إدخالها نتيجة لأسباب متعددة . إن آثار هذا النوع من الأخطاء خطيرة ولا تظهر مباشرة .
- استبدال وحدة بوحدة أخرى غير مدرجة ضمن قائمة أسماء الوحدات المختارة ، إذ قد يجد الباحث صعوبة فى جمع بيانات من وحدة فيأخذ وحدة أخرى (اختيار موظف عوضا عن الموظف المحدد بالعينة لعدم وجوده) .
- عدم التمكن من استكمال وصول جميع الاستمارات ، على الرغم من المتابعة المستمرة والزيارات المتكررة للوحدات ، خاصة إذا استخدمت طريقة المراسلة كطريقة لجمع البيانات .

وللتقليل من هذه الأخطاء المتعلقة بالتحيز فى الاختيار ، أو التخلص منها ، يمكننا اتباع ما يلى :

- اختيار جميع وحدات العينة عشوائيًا باستخدام إحدى طرق الاختيار العشوائى التى سنذكرها فيما بعد .
- عدم استبدال أية وحدة تم اختيارها بالعينة بوحدة أخرى .
- استكمال الإجابات لجميع الأسئلة ، واستلام جميع الاستمارات ، والقيام بالمتابعة المستمرة بالهاتف أو بالزيارات للعمل على استكمال استلام جميع الاستمارات .
- إجراء البحث التجريبى (العينة الاستطلاعية) لكشف التحيز المقصود وغير المقصود والتخلص منه أو التقليل من حجمه .
- تدريب الباحثين بشكل جيد على جمع البيانات والتقيد بالتعليمات المحددة المتعلقة بالوحدات المختارة .

ب - خطأ التحيز فى التقدير .

إضافة للأخطاء التى قد نقع فيها عند اختيار وحدات العينة ، هناك خطأ قد نقع فيه يتعلق بطريقة التقدير ، وطرق التحليل المستخدمة ، والتحيز الذى ينشأ بسبب عدم استخدام طرق التقدير أو التحليل المناسبة يسمى خطأ التحيز فى التقدير .

كمثال عن هذا النوع من الأخطاء نفترض أن لدينا ثلاث حيازات زراعية ، تزرع نوعًا من الخضراوات ، وكان متوسط محاصيلها على التوالى ١٥ ، ٢٠ ، ٢٥ طنًا للدونم الواحد .

إن تقدير متوسط المحصول للدونم بجمع المتوسطات الثلاثة وقسمتها على ثلاثة (أى $20 = \frac{15 + 20 + 25}{3}$) يعطى متوسط المحصول للدونم (٢٠ طناً) . إن هذه

الطريقة المستخدمة تؤدى إلى خطأ التحيز فى التقدير ، إذ يجب أن ترجح المتوسطات السابقة بالمساحات المزروعة فى المزارع الثلاث . فإذا كانت المساحات المزروعة فى هذه المزارع على التوالى هى ١٢ ، ٨ ، ١٤ دونماً فإن متوسط محصول الدونم يساوى :

$$C = \frac{(12 \times 15) + (8 \times 20) + (14 \times 25)}{12 + 8 + 14} = \frac{690}{34} = 20.29$$

ويكون مقدار خطأ التحيز فى التقدير (0.29 -) طن / دونم .

ويمكننا التقليل من أخطاء التحيز فى التقدير باستخدام طرق التقدير والتحليل المناسبة . ويجب على الإحصائى أن يكون حذراً فى استخدام تقدير متحيز . ويمكن تجاهل أثر التحيز إذا كان خطأ التحيز أقل من (٠,١٠) من الخطأ المعيارى ، ويعد التحيز الذى يكون ثابتاً من مجموعة لأخرى قليل الأهمية ، ولكن الأفضل التخلص منه أو التقليل منه باستخدام طرق التقدير والتحليل المناسبة * .

جـ - خطأ التحيز الناتج عن التعريف الخاطئ لوحدة المعاينة .

عندما نقوم بتحديد وحدة المعاينة ، يجب تعريفها تعريفًا واضحًا بشكل يقلل من أخطاء التحيز التى تنتج إذا كانت هذه الوحدة غير محددة وغير معرفة تعريفًا واضحًا . مثلاً ، عندما نحدد الموظف كوحدة إحصائية لجمع بيانات عن سنوات خبرته ومدى رضاه الوظيفى ، يجب أن نعرف الموظف تعريفًا واضحًا ، ويجب توضيح ما إذا كان الموظف المتعاقد الأجنبى سيعد من وحدات المعاينة ، وتبرز هذه المشكلة بشكل واضح عند اختيار وحدات لها مساحات أو قياسات معينة تختلف عن تلك التى يغطيها البحث ، وذلك بسبب عدم تعريفها تعريفًا واضحًا .

ويمكننا القول إن التحيز قد يكون مقصوداً أو غير مقصود سواء كان من قبل مصمم البحث أو العداد أو المدلى بالبيانات أو الذى يقوم بتحليلها ، ويجب التقليل منه بتصميم البحث بشكل جيد والتأكد من ذلك عن طريق العينة الاستطلاعية ، وتدريب الباحثين على جمع البيانات من الوحدات المختارة بدقة والقيام بالحملات الإعلامية للدعاية للبحث ، لكسب تعاون المدلين بالبيانات واستخدام طرق التحليل المناسبة .

---

* راجع الصيغة الرياضية للتحيز عند دراسة خواص التقدير الجيد فى الصفحات السابقة .

٢-٥-٣ الأخطاء الأخرى الشائعة فى العينات .

توجد أخطاء أخرى تقع عند استخدام المعاينة كأسلوب لجمع البيانات ، (وتقع أيضا عند استخدام أسلوب الحصر الشامل) نلخصها بما يلى :

- أخطاء عدم الاستجابة وقد تعود إلى عدم تحديث الإطار وشموله لجميع الوحدات ، أو عدم إمكانية الوصول إلى الوحدات المختارة ، أو عدم تواجد المدلين بالبيانات ، ويؤدى ذلك إلى زيادة أخطاء المعاينة العشوائية نتيجة انخفاض حجم العينة الفعالة (التى سنقوم بتحليل بياناتها) ، ويؤدى ذلك أيضًا إلى زيادة الأخطاء الأخرى .
- أخطاء التبويب ومعالجة البيانات ، وذلك بدءًا من تدقيق البيانات إلى عرضها بشكل جداول ويمكن التقليل من هذه الأخطاء عن طريق التدقيق وتصحيح الأخطاء .
- أخطاء الطباعة التى يجب تصحيحها .
- أخطاء تفسير النتائج على الرغم من صحة طرق التقدير وأساليب التحليل .

ويجب التقليل من جميع أنواع الأخطاء الأخرى للحصول على بيانات دقيقة تؤدى إلى نتائج سليمة .

الفصل
الثالث

المعاينة العشوائية البسيطة
Simple Random Sampling

٣ - ١ تعريف المعاينة العشوائية البسيطة :

تعد المعاينة العشوائية البسيطة أحد أنواع المعاينات الاحتمالية حيث تعتمد على نظرية الاحتمالات فى اختيار وحداتها وتقدير ثوابتها . وتعد هذه المعاينة أبسط أنواع المعاينات ، لكنها أهمها وأكثرها أصالة فى العشوائية .

تعرف المعاينة العشوائية بأنها طريقة اختيار (n) وحدة من مجتمع حجمه (N) بحيث يكون لكل عينة من العينات المكن اختيارها فرصة متساوية (احتمال متساو) فى الظهور .

ويتم اختيار وحدات المعاينة العشوائية البسيطة بحيث يعطى لكل وحدة من وحدات المعاينة الفرصة نفسها فى الظهور ، أى احتمال سحب أية وحدة متساوٍ عند اختيار كل وحدة من وحدات العينة . ولتوضيح التعريف السابق نورد المثال التالى :

إذا كان لدينا مجتمع من المصانع يتكون من (N) مصنعًا ونريد اختيار (n) مصنعًا لتقدير إنتاج ومبيعات وأرباح هذه المصانع وعدد العاملين فيها . لاستخراج عدد العينات المكن سحبها نميز بين حالتين :

أ - السحب مع عدم الإرجاع (عدم الإعادة) Selection without Replacement :

عند سحب الوحدة ، فإننا لا نعيد اختيارها مرة أخرى أى لا تعاد لتسحب ثانية . إن عدد العينات المكن سحبها فى حالة السحب مع عدم الإرجاع (عدم الإعادة) يساوى :

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!} \qquad \text{.... (3 - 1)}$$

حيث (!) تشير إلى عاملى العدد مثلاً !n تساوى :

$$n\,(n-1)\,(n-2)\,.......(3)\,(2)\,(1)$$

وعندما يكون احتمال ظهور أية عينة من هذه العينات المكن سحبها مساويًا إلى $\frac{1}{\binom{N}{n}}$ فإن المعاينة التى نحصل عليها تسمى معاينة عشوائية بسيطة [. ويلاحظ فى هذه الحالة أن احتمال اختيار أو ظهور أية وحدة متساوٍ فى كل مرة نسحب فيها ، ويساوى عند اختيار الوحدة الأولى $\frac{1}{(N)}$ إذ لكل وحدة من وحدات المعاينة الاحتمال نفسه للظهور ، ويساوى $\frac{1}{(N)}$. أما عند اختيار الوحدة الثانية فإن عدد وحدات المعاينة يساوى (N-1) لعدم إعادة الوحدة

التى تم اختيارها ويصبح احتمال ظهور أية وحدة فى السحب الثانى $\frac{1}{N-1}$ وفى السحب الثالث $\frac{1}{N-2}$ وهكـذا يكون احتمـال ظهور أيـة وحدة فى السحب الأخيـر $\frac{1}{N-(n-1)}$.

ونلاحظ فى هذه الحالة أن اختيار إحدى الوحدات تنفى ظهورها أكثر من مرة إذ تستبعد من عملية الاختيار فى السحوبات التالية .

ب – السحب مع الإرجاع (الإعادة) : Selection with Replacement

نعيد الوحدة التى يتم اختيارها ، أى فى حالة السحب المتكرر فإن عدد وحدات المعاينة التى يتكون منها المجتمع يبقى (N) وبالتالى فإن احتمـال ظهـور أية وحـدة فى كل سـحب يسـاوى $\frac{1}{(N)}$. أما عـدد العينات الممكن اختيارها فيساوى فى حالة السـحب مع الإرجاع (N^n) . ونلاحظ فى هذه الحالة أن اختـيار إحدى الوحـدات لا ينفى إعادة اختـيارها وتقاس الظاهرة حسب عدد مرات ظهور الوحدة فى العينة .

تطبيق (٣ – ١) :

بلغ عدد المصانع فى إحدى المناطق (٤) مصانع ، ونريد اختيار عينة حجمها مصنعان بأسلوب المعاينة العشوائية ، ما هو :

١ – عدد العينات الممكن سحبها إذا كان السحب مع عدم الإرجاع وما هى العينات الممكنة ؟

٢ – عدد العينات الممكن سحبها إذا كان السحب مع الإرجاع ؟

٣ – ما هو احتمال ظهور أية عينة ممكنة واحتمال ظهور الوحدة فى كلتا الحالتين ؟

الحـل :

١ – عدد العينات الممكن سحبها فى حالة السحب مع عدم الإرجاع :

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

$$\binom{4}{2} = \frac{4!}{2!\,(4-2)!} = \frac{4!}{2!\;2!}$$
$$= 6$$

أى أن عدد العينات الممكن اختيارها هو (٦) عينات .

إذا رمزنا إلى المصانع بالرموز أ ، ب ، ج ، د فإن العينات الممكنة هى :

العينة الأولى أ ، ب

العينة الثانية أ ، ج

العينة الثالثة أ ، د

العينة الرابعة ب ، ج

العينة الخامسة ب ، د

العينة السادسة ج ، د

والعينة التى تختارها هى إحدى العينات السابقة .

٢ – عدد العينات الممكن سحبها فى حالة السحب مع الإرجاع يساوى :

$$N^n = 4^2 = 16$$

أى عدد العينات الممكن سحبها هو (١٦) عينة .

٣ : أ – احتمال ظهور أى عينة ممكنة فى حالة السحب مع عدم الإرجاع :

$$\frac{1}{\binom{N}{n}} = \frac{1}{6}$$

وفى حالة السحب مع الإرجاع يساوى هذا الاحتمال :

$$\frac{1}{(N^n)} = \frac{1}{16}$$

ب – احتمال ظهور الوحدة فى حالة السحب مع عدم الإرجاع هو :

$$\frac{1}{4} , \frac{1}{3}$$

أما إذا كان السحب مع الإرجاع فيكون هذا الاحتمال مساوياً لـ : $\frac{1}{4} , \frac{1}{4}$ وهو متساوٍ فى كلا السحبين .

٣ - ٢ طرق اختيار العينة العشوائية البسيطة :

لعل السؤال الذى يتبادر إلى ذهن الباحث للوهلة الأولى هو : كيف يتم اختيار العينة العشوائية البسيطة بحيث يكون لكل عينة ممكنة حجمها (n) وحدة الفرصة نفسها (الاحتمال) فى الظهور ؟

إن الخطأ الكبير الذى يقع فيه كثير من الباحثين ، أن يقوم الباحث باختيار الوحدات بشكل كيفى ، أى وفق ما يراه الباحث مناسباً حيث يختار الوحدة وفق مزاجه ، ومن ثم يقول إن الاختيار كان عشوائيًا ، وإن العينة التى اختارها هى عينة عشوائية بسيطة . إن هذه العينة ليست عشوائية لاعتمادها على مزاج الباحث وأهوائه الشخصية .

ومن أجل اختيار العينة عشوائيًا ، يمكننا استخدام إحدى الطرق الخمس التالية :

- طريقة البطاقات .
- طريقة الكرات .
- طريقة عجلات الحظ .
- طريقة جداول الأرقام العشوائية .
- طريقة الاختيار العشوائى بالحاسوب .

وسنقوم باستعراض كل من الطرق السابقة باختصار مع التركيز على طريقة جداول الأرقام العشوائية لاستخدامها بشكل واسع فى المجالات العملية .

٣ - ٢ - ١ طريقة البطاقات :

تعد طريقة البطاقات إحدى طرق الاختيار العشوائى ، حيث نقوم بترقيم الوحدات الإحصائية بأرقام متسلسلة (الأرقام المتسلسلة للوحدات المدونة فى الإطار) ، ومن ثم تدون هذه الأرقام (وأحيانًا تدون الأسماء أيضًا) على بطاقات متشابهة تمامًا من حيث الشكل واللون والحجم . وتوضع البطاقات فى صندوق أو كيس وتخلط جيداً مع بعضها . ويتم اختيار عدد من البطاقات يساوى عدد وحدات العينة (حجم العينة) ، ويجب أن تخلط البطاقات جيداً بعد كل سحبة لضمان عشوائية الاختيار .

تتطلب هذه الطريقة جهوداً كبيرة من حيث إعداد وتجهيز البطاقات ، خاصة إذا كان حجم المجتمع كبيراً . لذا يفضل إذا كان حجم المجتمع كبيراً استخدام إحدى الطرق الأخرى لصعوبة تجهيز عدد كبير من البطاقات .

٣ - ٢ - ٢ طريقة الكرات :

يتم وفقًا لهذه الطريقة وضع الأرقام المتسلسلة داخل كرات متشابهة ومتجانسة من حيث الشكل واللون والحجم (أو كتابة الأرقام على الكرات) . ثم توضع هذه الكرات فى صندوق أو كيس وتخلط جيداً ويتم اختيار وحدات العينة المطلوبة ،ومن الضرورى خلط الكرات بعد كل سحبة لضمان عشوائية الاختيار .

٣ - ٢ - ٣ طريقة عجلات الحظ :

تتطلب هذه الطريقة ، تجهيز عدد من عجلات الحظ يساوى عدد منازل (خانات) حجم المجتمع . وترقم كل عجلة (عدا العجلة الأخيرة من اليسار) بالأرقام (٠ ، ١ ، ٢ ، . . . ، ٩) ، أما العجلة الأخيرة فترقم برقم المنزلة الأخيرة وما دون . فإذا كان حجم المجتمع (٣٠٠٠) موظف ونريد اختيار عينة بالأسلوب العشوائى ، فإننا نجهز أربع عجلات ترقم الثلاث الأولى من الصفر إلى (٩) . أما العجلة الأخيرة (المخصصة هنا للآلاف) فترقم بالأرقام (٠ ، ١ ، ٢ ، ٣) . ويتم اختبار هذه العجلات للتأكد من سلامتها وعدم تحيزها ، ومن ثم تختار وحدات العينة عن طريق تدوير العجلات مع بعض ، ويقرأ الرقم بجانب المؤشر ، فيكون رقم الوحدة الأولى المختارة ، وهكذا نكرر العملية عدداً من المرات إلى أن نحصل على العينة المطلوبة (مع ضرورة إهمال الرقم الذى يكون أكبر من حجم المجتمع) .

وتعد هذه الطريقة أسهل من الطريقتين السابقتين ، لكن استخدامها محدود إذا قورنت بطريقة جداول الأرقام العشوائية .

٣ - ٢ - ٤ طريقة جداول الأرقام العشوائية :

Tables of Random Numbers

تحتوى جداول الأرقام العشوائية على أعداد صحيحة تم إعدادها على أساس عشوائى وتقع بين الصفر والتسعة . وقد أدرجت هذه الأرقام فى صفحات ، يحتوى كل منها على عدد من الأسطر وعدد من الأعمدة (الحقول) ، كما هو موضح فى الملحق رقم (٤) ، وهناك نماذج متعددة أخرى من هذه الجداول ملحقة فى كثير من الكتب .

لتوضيح كيفية اختيار عينة عشوائية بسيطة باستخدام جداول الأرقام العشوائية ، نورد المثال التالى : لنفرض أننا نرغب اختيار عينة عشوائية بسيطة حجمها (٢٠٠) أسرة من أحد الأحياء الذى يحتوى على (٥٠٠٠) أسرة وذلك لتقدير متوسط عدد أفراد الأسرة .

لتحديد أرقام الأسر المختارة بالأسلوب العشوائى باستخدام جداول الأرقام العشوائية . نرقم الأسر بأرقام متسلسلة (0001, 0002, ..., 4999, 5000) (حيث نجد أن حجم المجتمع يساوى ٥٠٠٠ أسرة) . ويلاحظ أن عدد المنازل (الخانات) فى هذه الحالة أربع خانات ، لذا نحتاج إلى أربعة أعمدة . ونختار عشوائيًا أحد الجداول العشوائية ونختار برأس القلم أحد الأرقام عشوائيًا دون النظر إلى الجدول ونأخذ على يمينه عدداً من المنازل يساوى منازل المجتمع (أى أربعة أرقام فى المثال السابق بما فيه الرقم المختار) . ونبدأ بقراءة الأرقام من الأعلى إلى الأسفل بدءاً من النقطة المختارة وندون الأعداد التى تساوى حجم المجتمع أو أقل منه ، ونهمل الأعداد التى هى أكبر من حجم المجتمع . لنفرض أن نقطة البداية كانت نقطة تقاطع السطر الخامس مع العمود الثالث فى الصفحة الأولى أى الرقم (٢) ، نأخذ على يمينه ثلاثة منازل فيكون المختار الأول (٢١٢١) وهو رقم الأسرة الأولى المختارة ، ونقرأ الأعداد من الأعلى إلى الأسفل ، والرقم الثانى (٨٤٩٨) نهمله لأنه أكبر من (٥٠٠٠) والذى يليه (٣٥٢٣) هو رقم الأسرة الثانية المختارة ، وهكذا نتابع إلى أن نحصل على أرقام الأسر المختارة والتى تكون فى الأعمدة (٣ ، ٤ ، ٥ ، ٦) :

٢١٢١ ، ٣٥٢٣ ، ٠٨٧٩ ، ٢٧٥٣ ، ١٦٣٠ ، ٢٥١٠ ، ٣٩٧١ ، ٤٢٥٢ ، ٣٣٧٧ ، ٢٤٩٢ ، ٢٩٦٣ ، ٣٨٠٣ ، ٤٩٨٩ ، ..... ، ٤١٥٩ .

وعند الانتهاء نجد أننا نحتاج إلى أرقام مختارة أخرى ، لذا نترك العمود الثالث الذى بدأنا به ونضيف العمود السابع فتكون الأعمدة التى سنختار منها هى الرابع والخامس والسادس والسابع والتى تبدأ بالرقم (٥١٥٧) فنهمله ونهمل الذى يليه (٩٢٩٩) ونأخذ الرقم (٣٤٤٧) حيث يمثل رقم الأسرة المختارة ، ونكرر العملية نفسها إلى أن نحصل على أرقام وحدات العينة المختارة . ونلاحظ أن هناك مجموعات من الأرقام كثيرة تمكننا من اختيار أعداد كبيرة من الأرقام ، وعند الانتهاء من الصفحة نبدأ بالصفحة التى تليها بالأسلوب نفسه وذلك بإضافة العمود الأول من الصفحة الجديدة للأعمدة الثلاثة الأخيرة ثم أخذ عمودين من كل صفحة ... إلى أن نستخدم الأعمدة الأربعة الأولى من الصفحة الجديدة ، وهكذا نكرر العملية .

وقد نستخدم عدداً كبيراً من الصفحات لاختيار عينة ، وعند الانتهاء من جميع هذه الصفحات ، قد نصل إلى الأرقام نفسها التى بدأنا بها ، عندئذ نطرح حجم المجتمع من الرقم الذى نختاره ونحصل بذلك على أرقام جديدة . مثلاً إذا كان أحد الأرقام (٦٥٢٢) نطرح منه (٥٠٠٠) فيكون رقم الأسرة المختارة (١٥٢٢) وهكذا نتابع الاختيار . ويجب استخدام الرقم المختار مرة واحدة ، فإذا ظهر مرة أخرى نهمله .

تعد هذه الطريقة أسهل من الطرق السابقة ، ولا تتطلب سوى توفير جداول الأرقام العشوائية ، وترقيم وحدات المجتمع ، التى تكون غالباً مرقمة فى الإطار .

٣ - ٢ - ٥ الاختيار العشوائى بالحاسوب :

توجد برامج إحصائية جاهزة لتوليد الأرقام العشوائية باستخدام الحاسوب (الرئيسى والشخصى) حيث نحصل على قائمة بأرقام الوحدات المختارة ومن ثم نحصل على قائمة بأسماء وعناوين الوحدات المختارة وأهم المعلومات المتعلقة بها .

٣ - ٣ تقدير أهم معالم المجتمع :

٣ - ٣ - ١ رموز ومصطلحات :

سنستخدم الرموز والمصطلحات التالية لتقدير أهم معالم المجتمع :

- نرمز إلى حجم المجتمع بالرمز (N) وإلى عدد وحدات العينة التى نريد اختيارها أى حجم العينة بالرمز (n) .

- X_i تمثل الخاصية (أو الظاهرة) التى ندرسها للوحدة ذات الترتيب (i) فى المجتمع الإحصائى . ونجد أن قيم المجتمع أى مفردات المجتمع هى :

$$X_1, X_2, ..., X_N$$

ويكون مجموع قيم المجتمع ، ولنرمز له بالرمز (T) أو (X) :

$$X = T = X_1 + X_2 + ... + X_N$$

أى

$$T = \sum_{i=1}^{N} X_i \quad .... (3 - 2)$$

$$(i = 1, 2, ...N)$$

- أما متوسط المجتمع ($\overline{X}$) فيساوى :

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad .... (3 - 3)$$

ويرمز إلى متوسط المجتمع الحقيقى بالرمز (μ) .

- تباين المجتمع ولنرمز له بالرمز(σ^2) ويساوى إذا كان المجتمع محدوداً .

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2 \quad .... (3 - 4)$$

وتستخدم الصيغة التالية لتبسيط حسابه :

$$\sigma^2 = \frac{1}{N} \left[\sum_{i=1}^{N} X_i^2 - N \bar{X}^2 \right] \quad .... (3 - 5)$$

أما الانحراف المعيارى لقيم المجتمع (σ) فيساوى الجذر التربيعى لتباين المجتمع أى :

$$\sigma = \sqrt{\sigma^2} \quad .... (3 - 6)$$

ويرمز أحياناً لتباين قيم المجتمع بالرمز V (X)

كثيراً ما نستخدم صيغة أخرى لتباين قيم المجتمع عند دراسة التباين فى العينات ، وتسمى هذه الصيغة التباين المعدل للمجتمع المحدود (Adjusted Variance) :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \quad .... (3 - 7)$$

وتصبح الصيغتان S^2 و σ^2 متساويتين فى المجتمعات الكبيرة ، إذ يتقارب (N-1) مع (N) .

إن $S^2 = \frac{N}{N-1}\sigma^2$ (أو $\sigma^2 = \frac{N-1}{N} S^2$) ويصبح الكسر مساوياً للواحد الصحيح إذا كان حجم المجتمع كبيراً أى يصبح $S^2 = \sigma^2$.

٣ - ٢ - ٢ تقدير متوسط المجتمع والقيمة الكلية للمجتمع :

كثيراً ما نواجه حالات عملية يكون فيها متوسط قيم المجتمع مجهولاً ولا يمكن حسابه لعدم إمكانية القيام بحصر شامل ودقيق لوحدات المجتمع . لذا نلجأ إلى أسلوب المعاينة لتقدير أهم معلمات المجتمع من واقع بيانات العينة التى تم اختيارها من وحدات المجتمع بشكل عشوائى وتستخدم طريقتان لتقدير متوسط المجتمع والقيمة الكلية :

- التقدير بنقطة Point Estimation .
- التقدير بفترة Interval Estimation .

تقدير بنقطة لمتوسط المجتمع والقيمة الكلية للمجتمع :

(Mean and total Value Estimates)

يعد الوسط الحسابى لعينة عشوائية بسيطة مقدّراً غير متحيز للوسط الحسابى للمجتمع إذا رمزنا للقيمة (i) فى العينة بالرمز (x_i) يكون لدينا القيم التالية .

$$x_1 , x_2 , x_3 , \ldots , x_n$$

ويمكننا حساب الوسط الحسابى للعينة ($\bar{x}$) الذى هو تقدير غير متحيز لمتوسط المجتمع (μ) باستخدام الصيغة التالية* :

$$\boxed{\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i} \qquad . (3-8)$$

حيث نستخدم " μ " للدلالة على متوسط المجتمـع . إن الـوسط الحسـابى وفق الصيغة (3 - 8) هو مقدر غير متحيز لمتوسط المجتمع* . ونستخدم لتقدير القيمة الكلية للمجتمع (X) ولنزمر لها بالرمز (T) الصيغة التالية :

$$\boxed{\hat{T} = N \bar{x}} \qquad \ldots (3-9)$$

حيث $\hat{T} = \hat{X} = N \hat{\mu}$

ويعد هذا المقدر مقدراً غير متحيز للقيمة الكلية للمجتمع (T) .

---

* البرهان على ذلك ، تعلم أن

$$E(x_i) = \sum_{i=1}^{N} x_i P(x_i) = \sum_{i=1}^{N} x_i \frac{1}{N} = \bar{X} = \mu$$

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = \frac{1}{n} n\mu = \mu$$

أو بطريقة أخرى إذا كان عدد العينات الممكن سحبـها تساوي (NS)

$$E(\bar{x}) = \sum_{r=1}^{NS} \bar{x}_r P(\bar{x}_r) = \sum_{r=1}^{NS} \bar{x}_r \frac{1}{NS} = \mu$$

تطبيق (٣ – ٢) :

تتكون إحدى الإدارات من (١٠) موظفين كانت سنوات الخبرة لديهم كما يلى (بالسنوات) :

٦ ، ٤ ، ٥ ، ٣ ، ٥ ، ٦ ، ٧ ، ٦ ، ٣ ، ٥

المطلوب حساب :

١ – الوسط الحسابى لسنوات الخبرة للموظف وإجمالى سنوات الخبرة .

٢ – التباين والانحراف المعيارى لسنوات خبرة الموظف .

الحـــل :

البيانات السابقة تمثل قيم المجتمع ، لذا نستخدم الصيغ التالية :

– الوسط الحسابى لسنوات الخبرة للموظف :

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= \frac{1}{10} (6 + 4 + 5 + ... + 3 + 5)$$

$$= \frac{50}{10} = 5$$

أى أن متوسط سنوات الخبرة للموظف هو (٥) سنوات .

– إجمالى سنوات الخبرة لموظفى الإدارة :

$$X = T = \sum_{i=1}^{N} X_i = N \overline{X}$$

$$= (6 + 4 + 5 + ... + 3 + 5) = 50$$

أو يساوى

$$T = 10 x 5 = 50$$

أى أن إجمالى سنوات الخبرة للموظفين يساوى (٥٠) سنة .

– تباين قيم المجتمع والانحراف المعيارى :

$$V(X) = \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2$$

أو يساوى $V(X) = \sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{N} X_i^2 - N\,\overline{X}^2\right]$

$$= \frac{1}{10}\left[(6-5)^2 + (4-5)^2 + ... + (5-5)^2\right]$$

$$= \frac{1}{10}\left[1^2 + (-1)^2 + ... + (0)^2\right]$$

$$= \frac{16}{10} = 1.6$$

أو يساوى $\sigma^2 = \frac{1}{10}\left[266 - 10 \times 5^2\right]$

$$= \frac{266 - 250}{10} = \frac{16}{10} = 1.6$$

ويكون الانحراف المعيارى لسنوات خبرة الموظف :

$$\sigma = \sqrt{\sigma^2}$$

$$= \sqrt{1.6} = 1.265$$

أما التباين المعدل فيساوى :

$$S^2 = \frac{1}{N-1}\sum (X_i - \overline{X})^2 = \frac{16}{9} = 1.778$$

إن المقاييس السابقة تمثل مقاييس المجتمع ، وغالباً ما تكون مجهولة فى الحياة العملية خاصة فى المجتمعات الكبيرة ، لذا لابد من تقديرها من واقع بيانات عينة .

تطبيق (٣ – ٣) :

تتكون إحدى الوزارات من (١٠٠) موظف . نريد تقدير متوسط الإنفاق الشهرى للموظف وتقدير إجمالى إنفاقهم الشهرى . سحبنا عينة عشوائية بسيطة مكونة من (٥) موظفين ، فكان إنفاقهم الشهرى بالآلاف : ٤ ، ٦ ، ٥ ، ٧ ، ٨ .

ما هو تقدير الإنفاق الشهرى للموظف ومجموع إنفاق جميع الموظفين ؟

الحـــل :

– تقدير متوسط الإنفاق الشهرى للموظف :

$$\hat{X} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{5} (4 + 6 + 5 + 7 + 8)$$

$$= \frac{30}{5} = 6$$

أى (٦٠٠٠) ريال

– تقدير القيمة الكلية للإنفاق الشهرى :

$$\hat{X} = \hat{T} = N \bar{x}$$

$$= 100 \times 6 = 600$$

أى أن تقدير إجمالى إنفاق جميع الموظفين هو (٦٠٠) ألف ريال شهرياً .

٢ – ٣ – ٢ تباين التقديرات ومفرداتها :

تباين مفردات العينة : Variance of Sample Elements

كثيراً ما يستخدم تباين قيم العينة العشوائية البسيطة لتقدير تباين قيم المجتمع التى تكون مجهولة وذلك باستخدام الصيغة التالية :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \text{.... (3 – 10)}$$

إن تباين مفردات العينة (s^2) هو مقدر غير متحيز لتباين المجتمع المعدل S^2 ويمكن استخدام الرمز $V(x_i)$ للدلالة على تباين المفردة x_i المحسوبة من بيانات العينة أو $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = V(x_i) = s^2$$

ويكون الانحراف المعيارى لقيم العينة :

$$s = \sqrt{V(x_i)}$$

وعندما يكون حجم العينة (٣٠) فأكثر ، نضع فى المقام (n) عوضًا عن (n-1) إذ تتقارب القيمتان كلما أصبح حجم العينة كبيراً .

تباين تقدير متوسط المجتمع : (Variance of Mean Estimate)

تباين تقدير متوسط المجتمع $V(\bar{x})$ ويرمز له أحيانًا بالرمز $\sigma^2_{\bar{x}}$ ، ويتم حسابه باستخدام التوقع الرياضى :

$$\sigma^2_{\bar{x}} = V(\bar{x}) = E(\bar{x}-\mu)^2 \qquad .... (3-11)$$

ونميز هناك بين طريقتين للسحب .

أ – طريقة السحب مع الإعادة أو فى حالة المجتمع غير المحدود ، نجد فى هذه الحالة أن تباين تقدير متوسط المجتمع $V(\bar{x})$ والذى يسمى مربع الخطأ المعيارى (ويرمز له أحيانًا $\sigma^2_{\bar{x}}$) يساوى (كما هو موضح فى الملحق ٥ –١) :

$$V(\bar{x}) = \frac{\sigma^2}{n} \qquad .... (3-12)$$

حيث σ^2 هو تباين المجتمع . ويكون الخطأ المعيارى فى حالة السحب مع الإعادة أو فى حالة المجتمع غير المحدود $\sigma^2_{\bar{x}}$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad .... (3-13)$$

$$= \sqrt{V(\bar{x})}$$

ب – طريقة السحب مع عدم الإعادة أو إذا كان المجتمع محدوداً : ندخل فى هذه الحالة المعامل $\frac{N - n}{N - 1}$ على الصيغة (3-12) ويصبح تباين تقدير متوسط المجتمع فى هذه الحالة :

$$V(\bar{x}) = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1} = \frac{N - 1}{N} S^2 \frac{1}{n} \cdot \frac{N - n}{N - 1}$$

(حيث $\sigma^2 = \frac{N - 1}{N} S^2$) ، ونجد أن :

$$V(\bar{x}) = \frac{S^2}{n} \cdot \frac{N - n}{N} \qquad \text{.... (3 – 14)}$$

$$= \frac{S^2}{n} (1 - f)$$

حيث $f = \frac{n}{N}$ أى كسر المعاينة ، ويكون الخطأ المعيارى فى حالة السحب مع عدم الإعادة أو فى حالة المجتمع المحدود $\sigma^2_{\bar{x}}$:

$$\sigma^2_{\bar{x}} = \frac{S}{\sqrt{n}} \sqrt{1 - f} \qquad \text{.... (3 – 15)}$$

ولابد لنا من الإشارة إلى أنه يمكن إهمال كسر المعاينة إذا كان أقل من (٥٪) (وأحياناً إذا كان أقل من ١٠٪) . كما نستخدم تباين العينة (s^2) كمقدر غير متحيز لتباين المجتمع المعدل (S^2) إذا كان مجهولاً ، وذلك كما يتضح من الصيغ التالية :

- **تباين تقدير متوسط المجتمع باستخدام بيانات العينة :**

أ - فى حالة السحب مع الإعادة (أو المجتمع غير المحدود) :

لدينا :

$$Var(\bar{x}) = \frac{\sigma^2}{n}$$

وباستبدال σ^2 بما يساويها باستخدام S^2 نجد أن :

$$Var(\bar{x}) = \frac{S^2}{n} \times \frac{N-1}{N}$$

وكما نعلم فإن المقدر s^2 يعد مقدراً غير متحيز لتباين المجتمع S^2 فتصبح العلاقة السابقة باستخدام بيانات العينة :

$$\widehat{Var}(\bar{x}) = \frac{s^2}{n} \times \frac{N-1}{N} \qquad .... (3-16)$$

وعندما يكون حجم المجتمع كبيراً ، تتقارب N-1 مع N وتصبح العلاقة السابقة :

$$\widehat{Var}(\bar{x}) = \frac{s^2}{n} \qquad .... (3-17)$$

ويرمز لهذا التباين أحياناً بالرمز $(\hat{\sigma}^2_{\bar{x}})$ ، ويكون الخطأ المعيارى للتقدير :

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad .... (3-18)$$

ب - فى حالة السحب مع عدم الإعادة (أو المجتمع المحدود) :

نأخذ فى هذه الحالة بالاعتبار معامل تصحيح المجتمع المحدود ، وذلك باستخدام العلاقات السابقة ويصبح تباين تقدير متوسط المجتمع فى حالة السحب مع عدم الإعادة :

$$V(\bar{x}) = \frac{S^2}{n}(1-f)$$

ويصبح تباين تقدير المتوسط باستخدام بيانات العينة :

$$\hat{V}ar(\bar{x}) = \frac{s^2}{n}(1-f) \quad \text{.... (3 – 19)}$$

وعندما يكون حجم المجتمع كبيراً للغاية يصبح هذا التباين :

$$\hat{V}ar(\bar{x}) = \frac{s^2}{n} \quad \text{.... (3 – 20)}$$

ويكون الخطأ المعيارى للتقدير :

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{1-f} \quad \text{.... (3 – 21)}$$

ـ تباين تقدير القيمة الكلية للمجتمع :

نعلم أن تقدير القيمة الكلية للمجتمع ($\hat{X}$) يساوى :

$$\hat{X} = N\,\bar{x}$$

ويكون تباين تقدير القيمة الكلية للمجتمع ($\hat{X}$) يساوى :

$$V(\hat{x}) = \sigma^2_{\hat{x}} = V(N\bar{x})$$

$$= N^2 V(\bar{x})$$

ويساوى هذا التباين :

– فى حالة السحب مع عدم الإعادة :

$$V(\hat{x}) = N^2 \frac{\sigma^2}{n}\frac{N-n}{N-1} \quad \text{.... (3 – 22)}$$

وباستخدام التباين المعدل (S^2) نجد أن

$$V(\hat{x}) = \frac{N^2 S^2}{n} (1 - f) \quad .... (3 - 23)$$

– فى حالة السحب مع الإعادة :

$$V(\hat{x}) = N^2 \frac{\sigma^2}{n} \quad (3 - 24)$$

ويساوى هذا التباين باستخدام (S^2) :

$$V(\hat{x}) = \frac{N^2 S^2}{n} \quad \frac{N - 1}{N} \quad .... (3 - 25)$$

تطبيق (٣ – ٤) :

لدينا خمسة مرضى أعمارهم كما يلى :

١٠ ، ٢٠ ، ٥٠ ، ٤٠ ، ٣٠

المطلوب : استخراج :

١ – الوسط الحسابى لعمر المريض وإجمالى الأعمار .

٢.– تباين العمر للمجتمع والتباين المعدل والانحراف المعيارى .

الحـــل :

البيانات السابقة تمثل قيم المجتمع ، لذا نستخدم الصيغ المتعلقة بالمجتمع :

– الوسط الحسابى لعمر المريض :

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= \frac{1}{5}(10 + 20 + 50 + 40 + 30)$$

$$= \frac{150}{5} = 30$$

أى أن متوسط العمر هو (٣٠) سنة .

- القيمة الكلية للأعمار :

$$T = N \ \overline{X}$$

$$= 5 \times 30 = 150$$

أى أن مجموع الأعمار هو (١٥٠) سنة .

- تباين المجتمع :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{x})^2$$

$$= \frac{1}{5}\left[(-20)^2 + (-10)^2 + (20)^2 + (10)^2 + (0)^2\right]$$

$$= \frac{1000}{5} = 200$$

ويكون الانحراف المعيارى .

$$\sigma = \sqrt{\sigma^2}$$

$$= \sqrt{200} = 14.14$$

- التباين المعدل للمجتمع :

$$S^2 = \frac{1}{N-1}\sum(X_i - \overline{X})^2$$

$$= \frac{1000}{5-1} = \frac{1000}{4} = 250$$

ويكون الانحراف المعيارى المعدل :

$$S = \sqrt{250} = 15.81$$

تطبيق (٣ – ٥) :

سحبت عينة عشوائية بسيطة من مرضى إحدى المستشفيات لتقدير متوسط عمر المريض وكانت أعمار المرضى المختارين :

٣٠ ، ٤٠ ، ٥٠ ، ٤٠ ، ٦٠ ، ٢٠

المطلوب : استخراج :

١ – تقدير متوسط عمر المريض وتقدير إجمالى أعمار المرضى .

٢ – تباين العينة وتباين تقدير متوسط المجتمع من بيانات العينة ، إذا كان السحب مع الإعادة وإذا كان السحب بدون الإعادة .

(عدد مرضى المستشفى ٢٠٠ مريض وتباين المجتمع $\sigma^2 = 250$) .

الحـل :

– تقدير متوسط عمر المريض :

$$\hat{\bar{X}} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{6} (30 + 40 + ... + 20) = 40$$

أى أن متوسط العمر هو (٤٠) سنة .

– تقدير إجمالى أعمار المرضى :

$$\hat{\bar{X}} = \hat{T} = N \bar{x}$$

$$= 200 \times 40 = 8000$$

- تباين العينة :

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)$$

$$= \frac{1}{6-1}\,(10600 - 6 \times 40^2) = 200$$

ويكون الانحراف المعيارى للعينة :

$$s = \sqrt{s^2}$$

$$= \sqrt{200} = 14.14$$

- تباين تقدير متوسط المجتمع إذا كان السحب مع الإعادة :

نستخدم الصيغة التالية لأن تباين المجتمع معلوم :

$$\sigma^2_{\bar{x}} = V(\bar{x}) = \frac{\sigma^2}{n}$$

$$= \frac{250}{6} = 41.67$$

ويكون الخطأ المعيارى للتقدير :

$$\sigma_{\bar{x}} = \sqrt{V(\bar{x})}$$

$$= \sqrt{41.67} = 6.45$$

إذا افترضنا أن تباين المجتمع غير معلوم ، نستخدم الصيغة التالية :

$$\hat{V}(\bar{x}) = \hat{\sigma}^2_{\bar{x}} = \frac{N-1}{N}\,\frac{s^2}{n}$$

$$= \frac{200-1}{200} \times \frac{200}{6} = 33.17$$

ويكون الخطأ المعيارى لتقدير متوسط المجتمع :

$$\hat{\sigma}_{\bar{x}} = \sqrt{33.17} = 5.76$$

- تباين تقدير المتوسط إذا كان السحب مع عدم الإعادة :

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1} = \frac{S^2}{n} (1-f)$$

$$S^2 = \sigma^2 \frac{N}{N - 1} = 250 \times \frac{200}{199} = 251.25$$

$$\sigma_{\bar{x}}^2 = \frac{251.25}{6} (1- 0.03) = 40.62$$

والخطأ المعيارى للتقدير :

$$\sigma_{\bar{x}} = \sqrt{40.62} = 6.37$$

أما إذا استخدمنا تباين العينة فيكون تقدير تباين المتوسط :

$$\hat{\sigma}_{\bar{x}}^2 = \hat{V}(\bar{x}) = \frac{s^2}{n} (1 - f)$$

$$= \frac{200}{6} (1- \frac{6}{200}) = 32.33$$

والخطأ المعيارى للتقدير :

$$\hat{\sigma}_{\bar{x}} = \sqrt{32.33} = 5.69$$

٢ - ٢ - ٤ فترة الثقة لتقدير متوسط المجتمع وتقدير قيمته الكلية :

ذكرنا فيما سبق أن استخراج حدود الثقة لتقدير متوسط المجتمع يتطلب حساب تباين تقدير الوسط الحسابى (مربع الخطأ المعيارى للتقدير) الذى تختلف صيغته فى حالة سحب

عينة عشوائية بسيطة إذا كان السحب مع عدم الإعادة ، عنها في حالة السحب مع الإعادة ، ونميز بين الحالات التالية :

- يمكننا وضع حدى الثقة بمستوى ثقة $(1 - \alpha)\ \%$ إذا كان حجم العينة كبيراً (٣٠ فاكثر) :

$$\bar{x} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}\ (\bar{x})} \quad .... (3 - 26)$$

حيث نستخرج قيمة (Z) من جدول التوزيع الطبيعي بمستوى ثقة $(1 - \alpha)\ \%$ و $\hat{V}\ (\bar{x})$ عبارة عن مربع الخطأ المعياري للتقدير المحسوب من واقع بيانات العينة ويساوى :

$$\hat{V}\ (\bar{x}) = \frac{\hat{\sigma}^2}{n} \quad .... (3 - 27)$$

حيث

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

نظراً لأن حجم العينة كبير (٣٠ فاكثر) . وتستخدم الصيغة السابقة سواء كان السحب بدون إعادة أو مع الإعادة ، وذلك لأن كسر المعاينة $f = \frac{n}{N}$ يتلاشى بسبب صغر قيمته بالنسبة للمجتمع الكبير ، لذا نهمل $(1 - f)$ فى صيغة الخطأ المعيارى للتقدير ، سواء كان السحب مع الإعادة أو فى حالة السحب مع عدم الإعادة ، لأن قيمتها تساوى تقريباً الواحد الصحيح .

- أما إذا كان حجم العينة أقل من (٣٠) فإن حدى الثقة بمستوى ثقة $(1 - \alpha)\ \%$ هما :

$$\bar{x} \mp t_{(1-\alpha/2, n-1)} \sqrt{\hat{V}\ (\bar{x})} \quad .... (3 - 28)$$

حيث نستخرج قيمة (t) من جدول توزيع (ت) ستيودنت باحتمال $(1 - \alpha_{/2})$ وبدرجات حرية $(n - 1)$. أما $\hat{V}\ (\bar{x})$ فتساوى :

- إذا كان السحب مع الاعادة :

$$\hat{V}(\bar{x}) = \frac{s^2}{n} \frac{N-1}{N}$$

وإذا كان السحب مع عدم الإعادة :

$$\hat{V}(\bar{x}) = \frac{s^2}{n} (1-f)$$

حيـث :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

ولاستخراج حـدى الثقـة لتقديـر القيمـة الكليـة للمجتمـع ، نستخدم التبايـن المقـدر $\hat{V}(\hat{X})$ حيث :

$$\hat{V}(\hat{X}) = \hat{V}(N\bar{x})$$

$$= N^2 \hat{V}(\bar{x})$$

ويكون حدا الثقة إذا كان حجم العينة (٣٠) فاكثر :

$$\hat{X} \mp Z_{(1-\alpha/2)} \sqrt{\frac{N^2 s^2}{n}} \quad .... (3-29)$$

أما إذا كان حجم العينة أقل من (٣٠) ، فإن حدى الثقة هما :

$$\hat{X} \mp t_{(1-\alpha/2,\, n-1)} \sqrt{\frac{N^2 s^2}{n} (1-f)} \quad .... (3-30)$$

تطبيق (٣ – ٦) :

سحبت عينة عشوائية بسيطة حجمها (١٠٠) مريض لتقدير متوسط وزن المريض فى أحد المستشفيات التى تحتوى على (٢٠٠٠) مريض . إذا كان الوسط الحسابى للوزن (٦٠) كيلو غرامًا والانحراف المعيارى للعينة (٢٠) ، فما هى حدود الثقة للمتوسط وللقيمة الاجمالية بمستوى ثقة (٩٥٪) فى حالة السحب مع الإعادة وفى حالة السحب مع عدم الإعادة .

الحـــل :

نجد فى حالة السحب مع الاعادة أن :

– حدا الثقة للوسط الحسابى :

$$\overline{x} \mp Z_{(1-\alpha/2)} \frac{s}{\sqrt{n}}$$

حيث أهملنا $\frac{N-1}{N}$ لكبر حجم المجتمع ، أى أن حدى الثقة هما :

$$60 \mp 1.96 \frac{20}{\sqrt{100}}$$

$$60 \mp 3.92$$

وقد تم الحصول على قيمة (Z) بمستوى ثقة (٩٥٪) من اتجاهين من جداول منحنى التوزيع الطبيعى وتساوى (١,٩٦) . ويكون حدا الثقة ٥٦,٠٨ و ٦٣,٩٢ ويمكننا القول إن :

$$56.08 \leq \mu \leq 63.92$$

أى أنه لو سحبنا عدداً كبيراً من العينات العشوائية البسيطة ، حجم كل منها (١٠٠) مريض من مرضى المستشفى (مجتمع المرضى) وحسبنا حدود الثقة لكل عينة ، فإن (٩٥٪) من هذه الحدود لا بد أن تحتوى على متوسط المجتمع .

ويمكننا القول أيضًا إنه بمستوى ثقة (٠,٩٥) فإن متوسط وزن المريض فى المستشفى سيقع بين (٥٦,٠٨) سنة و (٦٣,٩٢) سنة .

– لتقدير القيمة الكلية لأوزان المرضى ، نستخدم الصيغة التالية .

$$\hat{X} \mp Z_{(1-\alpha/2)} \sqrt{\frac{N^2 \; \hat{\sigma}^2}{n}}$$

إن

$$\hat{X} = N \bar{x}$$

$$= 2000 \times 60 = 120\,000$$

وبالتالى يكون حد الثقة :

$$120\,000 \mp 1.96 \sqrt{\frac{(2000)^2 \quad (20)^2}{100}}$$

$$120\,000 \mp 7840$$

أى أن

$$112160 \leq X \leq 127840$$

أى بدرجة ثقة (٠.٩٥) فإن إجمالى أوزان المرضى X سيقع بين القيمتين (١١٢١٦٠) كيلو و (١٢٧٨٤٠) كيلو والتقدير بنقطة لهذا الإجمالى يساوى (١٢٠٠٠٠) .

أما فى حالة السحب مع عدم الإعادة ، فإننا فى الأحوال الاعتيادية نستخدم الصيغة التالية لاستخراج حدود الثقة للمتوسط :

$$\bar{x} \mp Z_{(1-\alpha/2)} \frac{s}{\sqrt{n}} \sqrt{1-f}$$

ونظراً لأن كسر المعاينة $f = \frac{n}{N} = \frac{100}{2000} = 0.05$ ، لذا يمكننا إهمال معامل تصحيح المجتمع المحدد (1-f) وبالتالى نحصل على النتائج السابقة نفسها.

تطبيق (٣ – ٧) :

لدراسة مستوى الإنفاق الشهرى لأحد الأحياء ، سحبت عينة عشوائية بسيطة حجمها (١٥) أسرة من أسر الحى التي عددها (٢٠٠٠) أسرة . وقد كان الإنفاق الشهرى للأسر من واقع بيانات العينة بالريالات كما يلي :

٤٠٠٠ ، ٤٠٠٠ ، ٥٠٠٠ ، ٢٠٠٠ ، ٣٠٠٠ ، ٢٠٠٠ ، ٣٠٠٠ ، ٣٠٠٠ ، ٤٠٠٠ ، ٣٠٠٠ ،
٢٠٠٠ ، ٣٠٠٠ ، ٤٠٠٠ ، ٢٠٠٠ ، ١٠٠٠

المطلوب :

تقدير المتوسط الشهرى لإنفاق الأسر وتقدير إجمالى الإنفاق فى هذا الحى بمستوى ثقة (٩٥٪) إذا كان السحب مع عدم الإعادة .

الحــل :

– تقدير متوسط الإنفاق الشهرى :

$$\hat{\bar{X}} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{4000 + 4000 + ... + 2000 + 1000}{15}$$

$$= 3000$$

أى (٣٠٠٠) ريال ، ويعد هذا التقدير غير متحيز لمتوسط إنفاق أسر الحى .

إن حدود الثقة بمستوى ثقة (٩٥٪) يساوى :

$$\bar{x} \mp t_{(1-\alpha/2,\, n-1)} \sqrt{\frac{s^2}{n}(1-f)}$$

لأن تباين المجتمع غير معلوم وتستخرج قيمة (t) من جداول ستيودنت ، إن

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

أو

$$= \frac{1}{n-1} \left(\sum x_i^2 - n\, \bar{x}^2\right)$$

$$= \frac{1}{15-1} \left\{\left[(4000)^2 + (4000)^2 + ... + (1000)^2\right] - (15 \times 3000^2)\right\}$$

$$= \frac{1}{14} \left[151\,000\,000 - (15 \times 9000\,000)\right]$$

$$= \frac{16000\,000}{14} = 1142857$$

ويكون الانحراف المعيارى :

$$s = \sqrt{1142857} = 1069.05$$

وبذلك تكون حدود الثقة بمستوى ثقة (٩٥٪) ودرجات حرية (n-1 = 14) :

$$\bar{x} \mp t_{(1-\alpha/2,\, n-1)} \sqrt{\frac{s^2}{n} \quad (1-f)}$$

$$3000 \mp 2.145 \sqrt{\frac{1142857}{15} \quad (1- \frac{15}{2000})}$$

$$= 3000 \mp (2.145 \times 274.99)$$

$$= 3000 \mp 589.85$$

أى أن :

$$2410.15 \leq \mu \leq 3589.85$$

ويمكننا القول إن الحد الأدنى للإنفاق هو (٢٤١٠.١٥) ريالات ، والحد الأعلى للإنفاق

(٣٥٨٩,٨٥) ريالاً وذلك بدرجة ثقة (٠,٩٥) ، وهذا يعنى أنه لو سحبنا عدداً كبيراً من العينات ذات الحجم (١٥) أسرة من المجتمع نفسه وحسبنا حدود الثقة لكل عينة ، فإن (٩٥٪) من هذه الحدود لا بد أن تحتوى على متوسط المجتمع .

- تقدير القيمة الكلية للإنفاق :

$$\hat{X} = \hat{T} = N \ \bar{x}$$

$$= 2000 \times 3000 = 6\ 000\ 000$$

وتكون حدود الثقة بمستوى ثقة (٩٥٪) :

$$\hat{X} \mp t_{(1-\alpha/2,\, n-1)} \sqrt{\frac{N^2\ s^2}{n}\ (1-f)}$$

$$= 6000\ 000 \mp (2.145 \times 549980)$$

$$= 6000\ 000 \mp 1179707$$

ويكون :

$$4820293 \leq \hat{X} \leq 7179707$$

ويمكن الحصول على نفس الحدين بضرب حدى المتوسط بحجم المجتمع أى بـ (٢٠٠٠) ، والفرق الصغير يعود للتقريب .

٣ - ٤ تقدير حجم العينة :

السؤال المهم الذى يطرحه الباحث : ما هو حجم العينة العشوائية البسيطة المناسب ؟ إن حجم العينة المناسب هو الذى نحدده لتقدير معالم المجتمع بدقة محددة ، وتتحدد هذه الدقة بدلالة الخطأ الذى يمكن قبوله عند تقدير المعالم والمخاطرة التى نقبل تحملها ، أى أن حجم العينة يتحدد بحيث يحقق خطأ ومخاطرة محددين .

إن حجم العينة الكبير يتطلب تكاليف مالية وبشرية ووقتًا كبيراً ، لكنه يعطى دقة أكبر ، وبالعكس فإن حجم العينة الصغير يؤدى إلى تكاليف مادية وبشرية ووقتًا أقل ، وقد تكون النتائج غير دقيقة . لذا فإن الأفضل تحديد حجم العينة على أساس دقة محددة مسبقًا .

٢ - ٤ - ١ حجم العينة لتقدير متوسط المجتمع :

إذا رمزنا للخطأ الذى نقبله فى تقدير متوسط المجتمع بالرمز (ß) وللمخاطرة أى احتمال الحصول على خطأ أكبر من (ß) التى نقبل تحملها بـ (α) فإننا نحدد حجم العينة بحيث يحقق :

$$P\left[\left|\bar{x} - \mu\right| \geq \beta\right] = \alpha$$

ونستطيع استخراج حجم العينة من حد الخطأ (ß) حيث يساوى هذا الحد لتقدير متوسط المجتمع μ :

أ - إذا كان المجتمع غير محدود أو السحب مع الإعادة :

$$\beta = Z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}$$

وبافتراض أن مستوى الثقة (٩٥٪) فـإن قيمـة $Z = 1.96 \approx 2$ تصبح قيمة ß :

$$\beta = 2 \frac{\sigma}{\sqrt{n}}$$

ونستخرج قيمة n من هذا المقدار فنجد أن :

$$\beta^2 = \frac{4\sigma^2}{n}$$

ومنه :

$$n = \frac{4\ \sigma^2}{\beta^2} \qquad \text{.... (3 – 31)}$$

وعند استخدام مستوى ثقة مختلف عن (٩٥٪) تصبح العلاقة السابقة :

$$n = \frac{Z^2\ \sigma^2}{\beta^2} \qquad \text{.... (3 – 32)}$$

حيث (Z) القيمة الجدولية المستخرجة من التوزيـع الطبيعـى المعيـارى بمسـتوى ثقة % (1 - α) و ß حد الخطأ الذى نقبله عند تقدير الوسط الحسابى . وفى حال عدم معرفة تباين

المجتمع σ^2 نقوم بتقديره من بيانات عينة استطلاعية (S^2 أو $\hat{\sigma}^2$) . ولتقدير حجم العينة بشكل نهائي ، لا بد من حساب كسر المعاينة $\frac{n}{N}$ فإذا كانت هذه النسبة أقل من (٠,٠٥) (وأحياناً إذا كانت أقل من ٠,١٠) فإننا نقبل حجم العينة المستخدم بالصيغة السابقة . أما إذا كان خلاف ذلك فإن حجم العينة النهائي يتحدد بالصيغة التالية بافتراض أن حجم العينة (المبدئي) الذى تم حسابه بالصيغة السابقة يساوي (n_0) إذا كان كسر المعاينة أكبر من (٠,٠٥) وأحياناً أكبر من (٠,١٠) :

$$n = \frac{n_0}{1 + \frac{no}{N}} \qquad .... (3-33)$$

ب – إذا كان المجتمع محدوداً أو فى حال السحب مع عدم الإعادة :

لا بد فى هذه الحالة من إدخال الكسر $\frac{N-n}{N-1}$

على حد خطأ التقدير المقبول ß فيصبح :

$$ß = Z_{(1-\alpha/2)} \sqrt{\frac{\sigma^2}{n} \frac{N-n}{N-1}}$$

$$ß^2 = Z^2 \frac{\sigma^2}{n} \frac{N-n}{N-1}$$

وبالقسمة على Z^2 والضرب في $n(N-1)$ والنقل واستخراج (n) خارج قوس نجد أن :

$$n\left[(N-1)\left(\frac{ß}{Z}\right)^2 + \sigma^2\right] = N\sigma^2$$

ومنه

$$n = \frac{N\sigma^2}{(N-1)\left(\frac{ß}{Z}\right)^2 + \sigma^2}$$

وبوضع $D = \left(\frac{B}{Z}\right)^2$ نجد أن :

$$n = \frac{N\ \sigma^2}{(N - 1)\ D + \sigma^2} \qquad .... (3 - 34)$$

وفى حالة عدم معرفة σ^2 نقدرها من بيانات عينة استطلاعية ونستخدم $\hat{\sigma}^2$ فى هذه الحالة أى تقدير تباين العينة . وإذا كان حجم العينة الاستطلاعية صغيراً نستخدم (s^2) حيث $\sigma^2 = \frac{N S^2}{N - 1}$ ونضع s^2 كمقدر غير متحيز للتباين المعدل S^2 .

وفى بعض الحالات فى حالة عدم معرفة σ^2 وعدم تقديره ، يتم تقديره باستخدام المدى وذلك باستخدام الصيغة التالية :

$$\hat{\sigma} \approx \frac{R}{4}$$

حيث R تمثل المدى المطلوب (الفرق بين أكبر قيمة وأصغر قيمة) ، ولكن الأفضل اختيار عينة استطلاعية لتقدير تباين المجتمع .

٣ - ٤ - ٢ حجم العينة لتقدير القيمة الكلية للمجتمع :

يصبح فى هذه الحالة حد خطأ التقدير B

$$B = Z\sqrt{V\,(N\,\bar{x})}$$

ونستخرج حجم العينة بالطرق السابقة نفسها حيث نجد أن :

$$n = \frac{N\ \sigma^2}{(N - 1)\ D + \sigma^2} \qquad .... (3 - 35)$$

حيث

$$D = \frac{B^2}{Z^2\ N^2}$$

تطبيق (٣ – ٨) :

نريد تقدير متوسط درجات عدد من الطلاب بخطأ تقدير (٣) درجات وبمستوى ثقة (٩٥٪) . إذا كان الانحراف المعيارى من واقع حصر شامل سابق يساوى (١٢) درجة . ما هو حجم العينة المناسب إذا كان إجمالى الطلاب (٢٠٠) طالب ، وكانت الطريقة المتبعة فى اختيار العينة السحب العشوائى مع الإعادة ؟

الحل :

حجم العينة يساوى فى حالة السحب مع الإعادة :

$$n = \frac{Z^2\ \sigma^2}{B^2}$$

$$= \frac{(1.96)^2\ (12)^2}{(3)^2} = 61.46 \approx 61$$

ونجد أن كسر المعاينة $f = \frac{61}{200} = 0.305$ وهى أكبر من (٥٪)

وأيضاً أكبر من (١٠٪) ، لذا يكون حجم العينة النهائى إذا كانت $(n_o = 61)$:

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

$$= \frac{61}{1 + \frac{61}{200}} = \frac{61}{1.305} = 46.74$$

$$\approx 47$$

أى أن حجم العينة المناسب هو (٤٧) طالباً .

تطبيق (٣ – ٩) :

نرغب فى تقدير متوسط وإجمالى الدخل الشهرى لأسر أحد الأحياء البالغ عددهم (١٠٠٠) أسرة وذلك بسحب عينة عشوائية بسيطة فإذا كان خطأ التقدير بحدود (٢٥) ريالاً ومعامل الثقة (٩٥٪) ، فالمطلوب :

١ – تحديد حجم العينة المناسب إذا سحبنا عينة استطلاعية حجمها (٤٠) أسرة وتم تقدير التباين $\sigma^2 = 5000$ ونريد تقدير متوسط الدخل .

٢ – تحديد حجم العينة إذا كان الفرق المتوقع بين أكبر دخل وأصغر دخل للأسرة (٢٠٠) ريال .

٣ – تحديد حجم العينة المناسب إذا رغبنا فى تقدير القيمة الكلية إذا كان الخطأ الذى نقبله فى إجمالى الدخل هو (٣٠٠٠٠) ريال .

(السحب مع عدم الإعادة) .

الحـل :

١ – حجم العينة المناسب لتقدير متوسط الدخل الشهرى يساوى :

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

حيث $D = \left(\frac{B}{Z}\right)^2$ ، أن $\sigma^2 = S^2$ لأن حجم المجتمع كبير ، لذا نضـع ($\hat{\sigma}^2$) ويكون :

$$n = \frac{1000 \times 5000}{(1000-1)\frac{(25)^2}{4} + 5000}$$

$$= \frac{5000\,000}{161093.75} = 31.04 = 31$$

ونجد أن كسر المعاينة $\frac{n}{N} = \frac{31}{1000} = 0.031$ أقل من (٠,٠٥) ، لذا يعد هذا الحجم نهائيًا .

٢ – إذا رغبنا أن نستخدم المدى لمتوسط الدخل ، نقوم بتقدير $\hat{\sigma}$ باستخدام الصيغة التالية :

$$\hat{\sigma} = \frac{R}{4} = \frac{200}{4} = 50$$

$$\hat{\sigma}^2 = 2500$$

ويكون حجم العينة :

$$n = \frac{1000 \times 2500}{(1000 - 1)\,\frac{(25)^2}{4} + 2500}$$

$$= \frac{2500\,000}{158593.75} = 15.76 \simeq 16$$

إن كسر المعاينة $f = \frac{n}{N} = \frac{16}{1000} = 0.016$ هو أقل من (٥٪)

لذا نعد هذا الحجم حجمًا نهائيًا أى أن حجم العينة المطلوب فى هذه الحالة هو (١٦) أسرة .

٣ – تقدير حجم العينة المناسب لتقدير القيمة الكلية :

$$n = \frac{N\,\sigma^2}{(N - 1)\,D + \sigma^2}$$

حيث :

$$D^2 = \frac{\beta^2}{Z^2\,N^2} = \frac{(30\,000)^2}{4 \times 1000\,000} = 225$$

ويكون حجم العينة :

$$n = \frac{1000 \times 5000}{(1000 - 1)\,(225) + 5000}$$

$$= 21.76 \simeq 22$$

وحيث إن كسر المعاينة أقل من (٠,٠٥) ، لذا يعد هذا العدد (٢٢) الحجم النهائى للعينة .

ويلاحظ اختلاف حجم العينة فى الطرق السابقة ، بسبب اختلاف الطرق المستخدمة .

الفصل
الرابع

معاينة نسبة المجتمع

Sampling of Population Proportion

٤ - ١ رموز وتعاريف :

كثيراً ما يهتم الباحث بتقدير نسبة المجتمع التي تتصف بصفة معينة ، مثلاً : قد يرغب أحد الباحثين في تقدير نسبة مشاهدى أحد البرامج التلفزيونية ، أو تقدير نسبة المتعطلين عن العمل ، أو تقدير نسبة المدخنين ، أو تقدير نسبة الموافقين على إجراء ما ، وأحياناً قد نريد تقدير نسبة الأشخاص الذين أعمارهم أكبر من (٦٠) سنة كما هو الحال في بحوث القوة العاملة بالعينة .

فى جميع هذه الحالات ، نرمز إلى مفردات المجتمع بالمتغير X حيث $(X_i = 1)$ إذا كانت المفردة (i) تتصف بالخاصية ، $(X_i = 0)$ إذا كانت المفردة (i) لا تتصف بالخاصية . وتلاحظ أن إجمالى عدد الذين يتصفون بالخاصية فى المجتمع :

$$T = X = \sum_{i=1}^{N} X_i \quad .... (4-1)$$

مثلاً عندما يكون الشخص (i) متعطلاً نضع $(X_i = 1)$ ، وعندما يكون غير متعطل نضع $(X_i = 0)$ ، ويكون إجمالى عدد المتعطلين يساوى (T) حسب الصيغة السابقة .

٤ - ٢ تقدير نسبة المجتمع والقيمة الكلية للمجتمع :

إن نسبة الذين يتصفون بالصفة أو الخاصية فى المجتمع (P) تساوى :

$$P = \frac{1}{N} \sum_{i=1}^{N} X_i \quad .... (4-2)$$

وتكون نسبة الذين لا يتصفون بالخاصية Q حيث :

$$P + Q = 1$$

أى

$$Q = 1 - P \quad .... (4-3)$$

إن نسبة المجتمع (P) غالباً ما تكون مجهولة ، لذا نقوم بتقديرها باستخدام أحد أنواع العينات .
إذا سحبنا عينة عشوائية بسيطة حجمها (n) وحدة ، تكون مفرداتها : $x_1, x_2, \ldots, x_n$

حيث $x_i = 1$ عندما تكون الوحدة متصفة بالخاصية و$x_i = 0$ عندما لا تتصف بالخاصية .
إن مقدر نسبة المجتمع للذين يتصفون بالخاصية يساوى :

$$p = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \text{.... (4 – 4)}$$

إن المقدر (p) هو مقدر غير متحيز لنسبة المجتمع (P) وذلك لأن :

$$E(p) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i)$$

$$= \frac{1}{n} \cdot n \cdot P = P$$

أما مقدر عدد الأفراد الذين يتصفون بالخاصية المدروسة ولنرمز له بالرمز $(\hat{T})$ فيساوى :

$$\hat{T} = \hat{X} = N\hat{P} = Np$$

أى

$$\hat{T} = Np \qquad \text{.... (4 – 5)}$$

ومقدر نسبة الذين لا يتصفون بالخاصية يساوى :

$$\hat{Q} = q = 1 - p$$

ومقدر إجمالى الذين لا يتصفون بالخاصية يساوى Nq أو (N - Np) .

تطبيق (٤ – ١) :

سحبنا عينة عشوائية بسيطة حجمها (٥٠) شخصاً من مجتمع عدد أفراده (١٠٠٠) شخص لتقدير نسبة المدخنين فى المجتمع . وقد وجدنا من بيانات العينة أن (٢٠) شخصاً يدخنون . ما هو تقدير نسبة المدخنين فى المجتمع وتقدير إجمالى عدد المدخنين ؟

الحـل :

– تقدير نسبة المدخنين :

$$p = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{50} (20) = 0.40$$

أى أن تقدير نسبة المدخنين (٤٠٪)

– تقدير إجمالى عدد المدخنين :

$$\hat{T} = \hat{X} = Np$$

$$= 1000 \times 0.40 = 400$$

– تقدير نسبة غير المدخنين :

$$\hat{Q} = q = 1 - p$$

$$= 1 - 0.4 = 0.6$$

أى (٦٠٪) ، وتقدير عدد غير المدخنين يساوى :

$$N - Np = 1000 - 400$$

$$= 600$$

٤ – ٢ تباين التقديرات لمعاينة النسب وتقديراتها :

٤ – ٢ – ١ تباين المجتمع وتباين العينة :

نعلم أن التباين المعدل للمفردة (X_i) يساوى :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})^2$$

أى

$$= \frac{1}{N-1} [\sum_{i=1}^{N} X_i^2 - N \overline{X}^2]$$

وبما أن X_i تساوى الواحد أو الصفر ، لذا فإن :

$$\sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} X_i = NP$$

حيث P نسبة المجتمع .

إن $\overline{X} = P$ و $\overline{X}^2 = P^2$ وبالتعويض نجد أن تباين المجتمع باستخدام نسبة المجتمع (P) :

$$S^2 = \frac{1}{N-1}(NP - NP^2)$$

$$= \frac{NP}{N-1}(1 - P)$$

أى

$$\boxed{S^2 = \frac{N}{N-1} P.Q} \qquad .... (4-6)$$

حيث $Q = 1 - P$.

كما ذكرنا سابقًا فإن نسبة المجتمع (P) غالبًا تكون مجهولة ، لذا نختار عينة عشوائية بسيطة ونجد (باستخدام الطريقة نفسها) أن تباين المفردة (x_i) باستخدام العينة العشوائية البسيطة :

$$\hat{S}^2 = s^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\,\overline{x}^2\right]$$

وحيث إن $\sum x_i^2 = \sum x_i = np$ ، لذا يكون لدينا :

$$s^2 = \frac{1}{n-1}(np - np^2)$$

أى

$$\boxed{s^2 = \frac{n}{n-1} p.q} \qquad .... (4-7)$$

ويعد (s^2) مقدرًا غير متحيز لـ S^2 .

٤ – ٣ – ٢ تباين تقدير نسبة المجتمع :

إذا رمزنا لتباين تقدير نسبة المجتمع V (p) ، يكون لدينا :

$$V(p) = E\ [p - E(p)]^2$$

$$= E\ [p - P]^2$$

ونميز بين حالتين لاستخراج تباين تقدير نسبة المجتمع :

أ – إذا كان السحب مع الإعادة أو فى حالة المجتمع غير المحدود :

نعلم أن :

$$V(\bar{x}) = \frac{S^2}{n}\ \frac{N-1}{N}$$

(حيث $\sigma^2 = S^2\ \frac{N-1}{N}$) ، وبتبديل قيمة (S^2) بقيمتها من الصيغة (4 - 6) نجد أن :

$$V(p) = \frac{PQ}{n}\ \frac{N}{N-1}\ \frac{N-1}{N} \qquad \text{.... (4 – 8)}$$

أى

$$V(p) = \frac{PQ}{n} \qquad \text{.... (4 – 9)}$$

ونهمل $\left(\frac{N}{N-1}\right)$ إذا كان حجم المجتمع كبيراً فى الصيغة (4 - 8) .

وعندما تكون نسبة المجتمع (P) مجهولة ، نقوم بتقديرها من بيانات عينة عشوائية بسيطة ، ويكون مقدر تباين نسبة المجتمع :

$$\hat{V}(p) = \frac{s^2}{n}\ \frac{N-1}{N}$$

$$= \frac{n}{n-1}\ \frac{1}{n}\ (pq)\ \frac{N-1}{N}$$

أى يساوى

$$\hat{V}(p) = \frac{pq}{n-1} \frac{N-1}{N} \qquad \text{.... (4 – 10)}$$

وعندما يكون حجم المجتمع كبيراً ، نهمل $\frac{N-1}{N}$ فيكون :

$$\hat{V}(p) = \frac{1}{n-1} pq \qquad \text{.... (4 – 11)}$$

ب – إذا كان السحب مع عدم الإعادة أو فى حالة المجتمع المحدود :

لدينا

$$V(p) = \frac{S^2}{n} \frac{N-n}{N}$$

وبتبديل قيمة $(S)^2$ بقيمتها من الصيغة (4 - 6) نجد أن :

$$V(p) = \frac{N}{N-1} \frac{PQ}{n} \frac{N-n}{N}$$

ومنه نجد أن :

$$V(p) = \frac{PQ}{n} \frac{N-n}{N-1} \qquad \text{.... (4 – 12)}$$

وعندما يكون حجم المجتمع كبيراً تتقارب (N-1) مع (N) وتصبح الصيغة السابقة :

$$V(p) = \frac{PQ}{n} (1-f) \qquad \text{.... (4 – 13)}$$

حيث $f = \frac{n}{N}$

وباستخدام بيانات عينة عشوائية بسيطة نجد أن مقدر تباين نسبة المجتمع يساوى :

$$\hat{V}(p) = \frac{pq}{n-1}\frac{N-n}{N-1} \quad .... (4-14)$$

إذا كان حجم المجتمع كبيراً :

$$\hat{V}(p) = \frac{pq}{n-1}(1-f) \quad .... (4-15)$$

ويكون الخطأ المعيارى لتقدير نسبة المجتمع من بيانات عينة :

$$\hat{\sigma}_P = \sqrt{\hat{V}(p)} \quad .... (4-16)$$

وذلك باستخدام إحدى الصيغتين الأخيرتين .

٤ - ٢ - ٣ تباين تقدير القيمة الكلية لمعاينة النسب :

نعلم أن مقدر القيمة الكلية لمعاينة النسب يساوى :

$$\hat{T} = \hat{X} = Np$$

ويكون تباين تقدير القيمة الكلية لمعاينة النسب باستخدام بيانات العينة :

$$\hat{V}(\hat{T}) = \hat{V}(Np)$$

$$= N^2\,\hat{V}(p)$$

وهكذا يمكننا استخدام الصيغ السابقة المتعلقة بتباين تقدير نسبة المجتمع وضربها فى (N^2) لنحصل على تباين تقدير القيمة الكلية المقدر من بيانات عينة .

– مقدر تباين تقدير القيمة الكلية لمعاينة النسب فى حال السحب مع الإعادة :

$$\hat{V}(\hat{T}) = N^2\frac{pq}{n-1}\frac{N-1}{N} \quad .... (4-17)$$

وعندما يكون حجم المجتمع كبيراً تتقارب (N-1) مع (N) وبالتالى تصبح الصيغة السابقة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} \qquad .... (4-18)$$

– تقدير تباين تقدير القيمة الكلية لمعاينة النسب فى حال السحب مع عدم الإعادة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} \frac{N-n}{N-1} \qquad .... (4-19)$$

وعندما يكون حجم المجتمع كبيراً تصبح الصيغة السابقة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} (1-f) \qquad .... (4-20)$$

حيث $f = \frac{n}{N}$

تطبيق (٤ – ٢) :

يتكون مجتمع من (٥) أفراد ، سحبنا منهم عينة عشوائية بسيطة حجمها شخصان (n =2) وذلك لتقدير نسبة المدخنين فى هذا المجتمع .

باستخدام الرمز $(X_i = 1)$ إذا كان الشخص يدخن ، والرمز $(X_i = 0)$ إذا كان **لا** يدخن وكانت قيمة (X_i) للأشخاص الخمسة :

الشخص	E	D	C	B	A
حالة التدخين (X_i)	0	1	0	1	1

المطلوب : استخراج :

١ – نسبة المدخنين ونسبة غير المدخنين فى المجتمع .

٢ – عدد العينات الممكن سحبها وماهية هذه العينات .

٣ – تقدير نسبة المدخنين وإثبات أنه تقدير غير متحيز لنسبة المدخنين فى المجتمع ، ثم وتقدير إجمالى المدخنين .

٤ – تقدير إجمالى عدد المدخنين وإجمالى عدد غير المدخنين .

٥ – تباين تقدير نسبة المجتمع باستخدام بيانات المجتمع ثم باستخدام بيانات العينة الثانية وذلك :

أ – فى حالة السحب بدون إعادة .

ب – فى حالة السحب مع الإعـادة .

الحــل :

١ – نسبة المدخنين فى المجتمع تساوى :

$$P = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= \frac{1}{5} (1 + 1 + 0 + 1 + 0) = \frac{3}{5} = 0.6$$

أى (٦٠٪) من الأشخاص يدخنون .

وتكون نسبة غير المدخنين فى المجتمع :

$$Q = 1 - P$$

$$= 1 - 0.60 = 0.40$$

أى (٤٠٪) من الأشخاص لا يدخنون .

٢ – عدد العينات الممكنة :

$$\binom{N}{n} = \frac{N!}{n!\,(N - n)!}$$

$$\binom{5}{2} = \frac{5!}{2!\,(5 - 2)!} = 10$$

أى هناك (١٠) عينات ممكنة ، والعينة التى نختارها باستخدام إحدى طرق السحب العشوائى هى إحداها . ويوضح الجدول التالى العينات الممكن سحبها وتقديرات نسبة المجتمع فيها .

رقم العينة	وحدات العينة	قيم العينة $x_1 \; x_2$	مجموع القيم $\sum x_i$	تقدير نسبة المجتمع $p = \frac{\sum x_i}{n}$	$q = 1-p$	$p \cdot q$
١	A, B	1,1	2	1.0	0.0	0.00
٢	A, C	1,0	1	0.5	0.5	0.25
٣	A, D	1,1	2	1.0	0.0	0.00
٤	A, E	1,0	1	0.5	0.5	0.25
٥	B, C	1,0	1	0.5	0.5	0.25
٦	B, D	1,1	2	1.0	0.0	0.00
٧	B, E	1,0	1	0.5	0.5	0.25
٨	C, D	0,1	1	0.5	0.5	0.25
٩	C, E	0,0	0	0.0	1.0	0.00
١٠	D, E	1,0	1	0.5	0.5	0.25
	المجموع					1.50

إن العينة التى نحصل عليها نتيجة السحب العشوائى هى إحدى العينات السابقة ، ولنفترض أنها العينة الثانية أى (A, C) .

٣ – يكون تقدير نسبة المدخنين فى المجتمع :

$$\hat{P} = p = \frac{1}{n} \sum x_i$$

$$= \frac{1}{2} (1 + 0) = 0.5$$

أما تقدير نسبة غير المدخنين فى المجتمع فتساوى :

$$\hat{Q} = q = 1 - p$$

$$= 1 - 0.5 = 0.5$$

إن (p) هو مقدر غير متحيز لنسبة المجتمع P وذلك لأن :

$$E(p) = E\left(\sum_{i=1}^{n} x_i / n\right)$$

إن احتمال سحب أى عينة ممكنة يساوى ($\frac{1}{10}$) ، لذا يكون :

$$E(p) = \frac{1}{2}[2 + 1 + 2 + ... + 1] \times \frac{1}{10}$$

$$= \frac{1}{2} \times 12 \times \frac{1}{10} = 0.6$$

وهى مساوية لنسبة المجتمع (P) . إذن (p) هو مقدر غير متحيز لنسبة المجتمع (P) ، لذا نقبل أية عينة ممكنة يتم سحبها تعد ممثلة للمجتمع الذى سحبت منه .

٤ - أما تقدير إجمالى عدد المدخنين فيساوى :

$$\hat{T} = \hat{X} = N\hat{P} = Np$$

$$= 5 \times 0.50 = 2.5$$

(وطبعًا فى هذه الحالة يمكن تقريبها إلى ٣ أشخاص)

وتقدير إجمالى غير المدخنين يساوى :

$$5 - 3 = 2$$

٥ - أ - تباين تقدير نسبة المجتمع باستخدام بيانات المجتمع :

- فى حال السحب مع الإعادة (مثالنا حجم المجتمع صغير ويساوى خمسة) :

$$V(p) = \frac{PQ}{n}$$

$$= \frac{(0.60) \times (0.40)}{2} = 0.12$$

ويكون الخطأ المعيارى للتقدير :

$$\sigma p = \sqrt{0.12} = 0.346$$

- فى حالة السحب مع عدم الإعادة :

$$V(p) = \frac{PQ}{n} \frac{N - n}{N - 1}$$

$$= \frac{0.60 \times 0.4}{2} \frac{5 - 2}{5 - 1}$$

$$= \frac{0.72}{8} = 0.09$$

ويكون الخطأ المعيارى للتقدير :

$$\sigma p = \sqrt{V(p)} = \sqrt{0.09} = 0.3$$

ب – تباين تقدير نسبة المجتمع باستخدام بيانات العينة :

– فى حال السحب مع الإعادة :

$$\hat{V}(p) = \frac{pq}{n-1} \cdot \frac{N-1}{N}$$

$$= \frac{0.5 \times 0.5}{2-1} \; \frac{5-1}{5} = \frac{1.0}{5}$$

$$= 0.20$$

ويكون الخطأ المعيارى للتقدير :

$$\hat{\sigma} p = \sqrt{\hat{V}(p)}$$

$$= \sqrt{0.20} = 0.447$$

– فى حالة السحب مع عدم الإعادة :

$$\hat{V}(p) = \frac{pq}{n-1} \; \frac{N-n}{N-1}$$

$$= \frac{0.5 \times 0.5}{2-1} \times \frac{5-2}{5-1} = \frac{0.75}{4}$$

$$= 0.187$$

ويكون الخطأ المعيارى للتقدير :

$$\hat{\sigma} p = 0.433$$

٤ - ٤ حدود الثقة لتقدير نسبة المجتمع وتقدير القيمة الكلية :

لاستخراج حدود الثقة لتقدير نسبة المجتمع ، نستخدم الأسلوب نفسه المستخدم عند استخراج حدود الثقة لتقدير متوسط المجتمع ، وذلك بعد إجراء التعديلات اللازمة ، وتصبح الصيغ المتعلقة بحدود الثقة لتقدير نسبة المجتمع باستخدام عينة عشوائية بسيطة :

$$p \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(p)} \qquad \text{.... } (4-21)$$

إذا كان حجم العينة كبيراً (٣٠ فاكثر) .

أما إذا كان حجم العينة أقل من (٣٠) فتصبح الصيغة السابقة :

$$p \mp t_{(1-\alpha/2, n-1)} \sqrt{\hat{V}(p)} \qquad \text{.... } (4-22)$$

حيث $\hat{V}(p)$ تساوى :

- فى حالة السحب مع الإعادة

$$\hat{V}(p) = \frac{pq}{n-1} \frac{N-1}{N} \qquad (4-23)$$

- فى حالة السحب مع عدم الإعادة :

$$\hat{V}(p) = \frac{pq}{n-1} \frac{N-n}{N} \qquad \text{... } (4-24)$$

وذلك إذا كان حجم المجتمع كبيراً ، ونضع (N-1) عوضاً عن (N) إذا كان حجم المجتمع صغيراً . أما حدود الثقة لتقدير القيمة الكلية لمعاينة النسب فهى :

$$\hat{T} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{T})} \qquad \text{... } (4-25)$$

أو

$$\hat{T} \mp t_{(1-\alpha/2, n-1)} \sqrt{\hat{V}(\hat{T})} \qquad \text{.... } (4-26)$$

حيث $\hat{V}(\hat{T})$ تساوى :

أ – فى حالة السحب مع الإعادة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} \frac{N-1}{N} \quad \text{.... (4 – 27)}$$

وعندما يكون حجم المجتمع كبيراً تصبح هذه الصيغة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} \quad \text{... (4 – 28)}$$

ب – فى حالة السحب مع عدم الإعادة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} \frac{N-n}{N-1} \quad \text{... (4 – 29)}$$

وإذا كان حجم المجتمع كبيراً تصبح هذه الصيغة :

$$\hat{V}(\hat{T}) = N^2 \frac{pq}{n-1} (1-f) \quad \text{.... (4 – 30)}$$

حيث $f = \frac{n}{N}$

تطبيق (٤–٣) :

يتكون مجتمع من (٢٠٠٠) موظف يعملون فى إحدى الوزارات . سحبت عينة عشوائية بسيطة حجمها (١٠٠) موظف لمعرفة آرائهم حول الإجراءات الجديدة التى طبقت فى هذه الوزارة ، وقد تبين أن (٦٠) موظفًا يرون أن هذه الإجراءات فعالة .

ما هو تقدير نسبة الذين يرون أن الإجراءات الجديدة فعالة ؟ وما هو تقدير إجمالى الموظفين الذين يرون ذلك بمستوى ثقة (٩٥٪) ؟ (السحب مع عدم الأعادة) .

الحـل :

أ – تقدير نسبة الذين يرون أن الإجراءات فعالة :

$$\hat{P} = p = \frac{\sum x_i}{n}$$

$$= \frac{60}{100} = 0.60$$

ب – ويكون تقدير نسبة الذين لا يرون ذلك :

$$\hat{Q} = q = 1 - p$$

$$= 1 - 0.60 = 0.40$$

وتكون حدود الثقة :

$$p \mp Z_{(1-\alpha/2)} \sqrt{\frac{pq}{n-1}(1-f)}$$

$$0.6 \mp 1.96 \sqrt{\frac{0.6 \times 0.4}{100-1}\left(1 - \frac{100}{2000}\right)}$$

$$0.6 \mp 0.094$$

ويكون الحد الأدنى لتقدير النسبة :

$$0.6 - 0.094 = 0.506$$

والحد الأعلى :

$$0.6 + 0.094 = 0.694$$

أي أن :

$$0.506 \leq P \leq 0.694$$

أي أن نسبة الذين يرون أن الإجراءات فعالة في هذه الوزارة ستتراوح بين (٠,٥٠٦) و(٠,٦٩٤) بدرجة ثقة (٩٥٪) ، ويمكننا القول إننا لو سحبنا عدداً كبيراً من العينات العشوائية البسيطة ذات الحجم (١٠٠) من المجتمع نفسه ، وحسبنا حدود الثقة لهذه العينات لنسبة الذين يرون أن الإجراءات فعالة ، فإننا نتوقع أن (٩٥٪) من هذه الحدود تتضمن نسبة المجتمع (أي نسبة الذين يرون أن الإجراءات فعالة في الوزارة) .

ب – تقدير إجمالى الموظفين الذين يرون أن الإجراءات فعالة :

$$\hat{T} = N\hat{p}$$

$$= 2000 \times 0.6 = 1200$$

وتكون حدود الثقة :

$$\hat{T} \mp Z_{(1-\alpha/2)} \sqrt{N^2 \frac{pq}{n-1} (1-f)}$$

$$1200 \mp 1.96 \sqrt{(2000)^2 \frac{0.6 \times 0.4}{100-1} \left(1 - \frac{100}{2000}\right)}$$

$$1200 \mp 188$$

ويكون حدا الثقة بمستوى ثقة (٩٥٪) :

$$1012 \leq T \leq 1388$$

أى أن إجمالى الموظفين الذين يرون أن الإجراءات المطبقة فعالة سيقع بين ١٠١٢ و ١٣٨٨ موظفًا ، وذلك بمستوى ثقة (٩٥٪) ، كما يمكننا القول إننا لو سحبنا عددًا كبيرًا من العينات العشوائية البسيطة من المجتمع نفسه ولها نفس الحجم ، وحسبنا حدود الثقة للذين يرون أن الإجراءات فعالة ، فإن (٩٥٪) من حدود الثقة لهذه العينات ستتضمن إجمالى الذين يرون ذلك فى هذه الوزارة .

٤ – ه تحديد حجم العينة فى معاينة النسب :

لتحديد حجم العينة فى معاينة النسب ، نستخدم الصيغ المستخدمة عند تحديد حجم العينة لتقديرى متوسط المجتمع والقيمة الكلية للمجتمع ، مع استبدال σ^2 بما يساويها حيث $\sigma^2 = PQ$ ، وتصبح الصيغ السابقة كما يلى :

٤ – ه –١ حجم العينة لتقدير نسبة المجتمع :

إذا كان السحب مع الإعادة :

$$n = Z^2 \frac{PQ}{B^2} \quad .... (4-31)$$

حيث (β) هو حد الخطأ الذى نقبله عند تقدير نسبة المجتمع . ويكون هذا الحجم نهائيًا إذا كان كسر المعاينة أصغر من (٠,٠٥) (وأحيانًا إذا كان أصغر من ٠,١٠) . أما إذا كان كسر المعاينة أكبر من (٠,٠٥) (أو ٠,١٠) فيصبح هذا الحجم مبدئيًا ويساوى (n_o) ويكون الحجم النهائى للعينة :

$$n = \frac{n_o}{1 + \frac{n_o}{N}} \quad \text{.... } (4-32)$$

– إذا كان السحب مع عدم الإعادة :

$$n = \frac{N\,P\,Q}{(N-1)\,D + PQ} \quad \text{.... } (4-33)$$

حيث $D = \frac{\beta^2}{Z^2}$

٤ – ٥ – ٢ حجم العينة لتقدير القيمة الكلية فى معاينة النسب :

نستخدم الصيغ السابقة ولكن نضع فى هذه الحالة :

$$D = \frac{\beta^2}{Z^2 N^2}$$

وتقدر P من عينة استطلاعية فى جميع الحالات السابقة أو من بحث سابق أو باستخدام طريقة المدى الموضحة فيما سبق :

$$\hat{\sigma} = \sqrt{PQ} \approx \frac{R}{4}$$

تطبيق (٤ – ٤) :

ترغب إحدى المؤسسات فى سحب عينة عشوائية بسيطة لتقدير نسبة المستهلكين الذين يرون أن الإنتاج مناسب من حيث الجودة ، وذلك بخطأ تقدير (٠,٠٥) . وقد تبين من بحث

سابق أن (٠.٥٠) من المستهلكين يرون أن الإنتاج مناسب . ما هو حجم العينة المناسب لتقدير النسبة (بمستوى ثقة ٩٥٪) إذا كان عدد المستهلكين لإنتاج المؤسسة (٢٠٠٠) مستهلك (حالة السحب مع عدم الإعادة و حالة السحب مع الإعادة) .

الحل :

أ - حالة السحب مع عدم الإعادة :

$$n = \frac{N\,P\,Q}{(N-1)\,D + PQ}$$

$$D = \frac{B^2}{Z^2} = \frac{(0.05)^2}{(1.96)^2} = \frac{0.0025}{3.8416}$$

$$= 0.00065$$

$$n = \frac{2000 \times 0.5 \times 0.5}{(2000-1) \times 0.00065 + (0.5 \times 0.5)}$$

$$= \frac{500}{1.54935} = 322.7 \approx 323$$

أي أن حجم العينة المناسب هو (٣٢٣) مستهلكًا .

ب - حالة السحب مع الإعادة :

$$n = \frac{Z^2\,P\,Q}{B^2}$$

$$= \frac{(1.96)^2 \times 0.5 \times 0.5}{(0.05)^2} = \frac{0.9604}{0.0025}$$

$$= 384.16 \approx 384$$

في هذه الحالة نجد أن كسر المعاينة يساوي :

$$f = \frac{n}{N} = \frac{384}{2000} = 0.192$$

ونظراً لأن كسـر المعاينة أكبر من (٠,٠٥) وأيضًا أكبر من (٠,١٠) لذا يعد الحجم السـابق مبدئيًا ($n_o = 384$) ويكون حجم العينة النهائى :

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

$$n = \frac{384}{1 + \frac{384}{2000}} = \frac{384}{1.192} = 322.14$$

$$= 322$$

وهو حجم العينة النهائى لتقدير نسبة المجتمع بمستوى ثقة (٩٥٪) وخطأ تقدير (٠,٠٥) .

الفصل
الخامس

المعاينة الطبقية العشوائية
Random Stratified Sampling

٥ - ١ تعريف المعاينة الطبقية العشوائية :

نستطيع فى بعض الأحيان ، تقسيم المجتمع الذى نقوم بدراسته إلى أقسام (أو طبقات Stratas) مختلفة فيما بينها من حيث الخاصية التى نقيسها ، بينما نجد أن هناك تشابهاً بين مفردات كل طبقة أكثر من تشابه المفردات داخل المجتمع بأكمله . وعند استخدام المعاينة الطبقية تكون التباينات بين مفردات كل طبقة أقل من التباينات الموجودة بين الطبقات .

ويتم تقسيم المجتمع إلى طبقات باستخدام عدة أسس ، مثلاً يمكن تقسيم إحدى الدول على أساس جغرافى إلى عدة مناطق جغرافية (مدن ، مناطق ، محافظات) يسمى كل منها طبقة . كما يمكن تقسيم المجتمع على أساس نوعى كتقسيم المصانع حسب نوع الصناعة (طبقة الصناعات الغذائية ، طبقة الصناعات النسيجية ، ...) أو حسب حجم المصنع من حيث الإنتاج وعدد العاملين (طبقة المصانع الكبيرة ، طبقة المصانع المتوسطة ، طبقة المصانع الصغيرة) .

ويمكننا تعريف المعاينة الطبقية العشوائية بأنها عملية اختيار عدد من الوحدات من مجتمع مقسم إلى طبقات (بحيث تكون الطبقات غير متداخلة وتكون المفردات ضمن الطبقة الواحدة متجانسة ، بينما هناك فروق كبيرة بين الطبقات) ، ويتم اختيار عينة عشوائية من كل طبقة بحيث يكون السحب من الطبقات المختلفة مستقلاً ، ومجموع العينات المختارة من الطبقات تشكل العينة الطبقية العشوائية ، وذلك للوصول إلى خصائص المجتمع من بيانات هذه العينة . إننا نعدّ كل طبقة مجتمعاً صغيراً ، تسحب منه عشوائياً عينة ذات حجم محدد ، ونقوم بتقدير معالم المجتمع لكل طبقة على حدة ، ثم تستخدم هذه التقديرات لتقدير معالم المجتمع كلّه .

إن الخطأ المعيارى للعينة يتأثر بشكل عام بتشتت مفردات المجتمع الذى سحبت منه ، لذا نجد أن الخطأ المعيارى للعينة الطبقية ، أقل من الخطأ المعيارى للعينة العشوائية البسيطة ، نتيجة لإزالة قسم من تشتت المجتمع الإحصائى بإلغاء الاختلافات الكبيرة الموجودة ضمن الطبقة الواحدة .

ويتساوى الخطأ المعيارى من كلتا العينتين إذا كان المجتمع متجانساً تماماً ، وهكذا نجد أن التقديرات التى يصل إليها الباحث باستخدام المعاينة الطبقية ، أكثر دقة من التقديرات التى يتوصل إليها باستخدام العينة العشوائية البسيطة .

وتستخدم المعاينات الطبقية بشكل واسع فى البحوث المختلفة للأسباب التالية :

- الحصول على بيانات تفصيلية عن كل طبقة من طبقات المجتمع .
- الحصول على تقديرات أكثر دقة إذا كان هناك اختلاف ملحوظ بين مفردات المجتمع من حيث الخاصية التى ندرسها . مثلاً ، عند دراسة مستوى الدخل ، نجد أن هناك اختلافاً

كبيراً بين دخول الأفراد ، لذا نقسمهم إلى ثلاث طبقات : أصحاب الدخول المرتفعة ، أصحاب الدخول المتوسطة ، أصحاب الدخول المنخفضة . والتقديرات التي نحصل عليها تكون أكثر دقة من غيرها .

- الحصول على تقديرات لكل طبقة ومن ثم تقديرات لمعالم المجتمع كلّه .
- نستطيع إدخال عنصر التكاليف المتعلقة بجمع البيانات وتبويبها عند تحديد حجم كل طبقة ، خاصة إذا كانت التكاليف تختلف من طبقة لأخرى بشكل كبير .
- تعدّ المعاينة الطبقية مناسبة أكثر من غيرها من المعاينات وذات أثر فعال إذا كان المجتمع يتضمن قيماً متطرفة لأننا نستطيع جمعها في طبقة واحدة .
- ولا بد لنا من الإشارة إلى أن المعاينة الطبقية تتشابه مع المعاينة العشوائية البسيطة في أن كلا النوعين ، يعدّان من العينات الاحتمالية ، حيث يكون لكل وحدة في المجتمع فرصة احتمالية محددة للاختيار في العينة . كما أن مقدرات كلتا الطريقتين هي مقدرات غير متحيزة ومتسقة لأنه يمكن الحصول على تقديرات قيم معالم المجتمع من نتائج العينة .

٥ - ٢ رموز وتعاريف :

- إذا استخدمنا الرموز التالية :

N عدد وحدات المجتمع .

L عدد الطبقات التي ينقسم إليها المجتمع .

N_h عدد وحدات الطبقة ذات الرتبة (h) .

n حجم العينة الطبقية .

n_h حجم العينة المسحوبة من الطبقة ذات الرتبة (h) .

نجد أن حجم المجتمع الإجمالي قد قسم إلى مجتمعات صغيرة عددها (L) طبقة وهي :

$N_1, N_2, N_3, ..., N_L$

وحجم المجتمع يساوي :

$$N = N_1 + N_2 + ... + N_L$$

أي أن :

$$N = \sum_{h=1}^{L} N_h \qquad .... (5 - 1)$$

(حيث $h = 1,2,3, ...L$)

كذلك نجد أن العينة التى تم سحبها من جميع الطبقات تتكون من عينات جزئية عددها (L) عينة وهى :

$n_1, n_2, n_3, ..., n_L$

أى أن حجم العينة الطبقية يساوى :

$n = n_1 + n_2 + n_3 + ... + n_L$

أى أن :

$$n = \sum_{h=1}^{L} n_h \qquad .... (5-2)$$

– إذا رمزنا إلى قيمة الخاصية فى المجتمع للوحدة (i) فى الطبقة (h) بالرمز (X_{hi}) ، فإننا نجد أن مجموع قيم المفردات الموجودة فى الطبقة (h) فى المجتمع ولنرمز له بالرمز (X_h) يساوى :

$X_h = X_{h1} + X_{h2} + ... + X_{hN_h}$

أى أن :

$$X_h = \sum_{i=1}^{N_h} X_{hi} \qquad .... (5-3)$$

وهكذا نجد أن مجموع مفردات الطبقة الأولى فى المجتمع التى حجمها (N_1) هى :

$X_1 = X_{11} + X_{12} + ... + X_{1N_1}$

$$= \sum_{i=1}^{N_1} X_{1i}$$

ومجموع مفردات الطبقة الثانية :

$$X_2 = \sum_{i=1}^{N_2} X_{2i}$$

ومجموع مفردات الطبقة (h) يساوى فى المجتمع :

$$X_h = \sum_{i=1}^{N_h} X_{hi} \qquad \text{.... (5 - 4)}$$

وحيث يوجد لدينا (L) طبقة ، فإن مجموع قيم مفردات المجتمع ولنرمز له بالرمز (X) يساوى :

$$X = X_1 + X_2 + ... + X_L$$

$$= \sum_{h=1}^{L} X_h$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}$$

وذلك بتبديل (X_h) بقيمتها ، أى أن :

$$X = \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi} \qquad \text{.... (5 - 5)}$$

– كذلك إذا رمزنا إلى مجموع قيم مفردات العينة من الطبقة (h) بالرمز (x_h) ، نجد أن :

$$x_h = \sum_{i=1}^{n_h} x_{hi} \qquad \text{.... (5 - 6)}$$

وهكذا نجد أن مجموع قيم مفردات العينة من الطبقة الأولى تساوى :

$$x_1 = x_{11} + x_{12} + x_{13} ..... + x_{1n_1}$$

ومجموع قيم مفردات العينة من الطبقة الثانية هو :

$$x_2 = x_{21} + x_{22} + x_{23} ..... + x_{2n_2}$$

ومجموع قيم مفردات العينة من الطبقة (h) يساوى :

$$x_h = x_{h1} + x_{h2} + x_{h3} ..... + x_{hn_h}$$

أى تساوى :

$$x_h = \sum_{i=1}^{n_h} x_{hi}$$

وتكون القيمة الكلية لمفردات العينة من جميع الطبقات ولنرمز لها بالرمز (x) تساوى :

$$x = x_1 + x_2 + x_3 + \ldots + x_L$$

وبالتالى يكون :

$$x = \sum_{h=1}^{L} x_h$$

أى مجموع قيم مفردات العينة الطبقية يساوى :

$$x = \sum_{h=1}^{L} \sum_{i=1}^{n_h} X_{hi} \qquad \ldots (5-7)$$

– متوسط الطبقة (h) فى المجتمع ولنرمز له بالرمز ($\overline{X}_h$) يساوى :

$$\overline{X}_h = \frac{X_h}{N_h}$$

أى يساوى :

$$\overline{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi} \qquad \ldots (5-8)$$

ومتوسط الطبقة (h) فى العينة ولنرمز له بالرمز ($\overline{x}_h$) يساوى :

$$\overline{x}_h = \frac{x_h}{n_h}$$

أى يساوى :

$$\overline{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \qquad \ldots (5-9)$$

- المتوسط العام للمجتمع ولنرمز له بالرمز ($\overline{X}$) يساوى :

$$\overline{X} = \frac{X}{N}$$

$$\overline{X} = \frac{\sum_{h=1}^{L} N_h \overline{X}_h}{N} \qquad .... (5 - 10)$$

أى أن المتوسط العام للمجتمع يساوى مجموع متوسطات الطبقات فى المجتمع مرجحة بنسبة عدد وحدات كل طبقة إلى عدد وحدات المجتمع . فإذا رمزنا إلى نسبة عدد وحدات الطبقة (h) (N_h) وإلى عدد وحدات المجتمع (N) بالرمز (W_h) يكون :

$$W_h = \frac{N_h}{N}$$

وبالتالى يصبح المتوسط العام للمجتمع :

$$\overline{X} = W_1 \overline{X}_1 + W_2 \overline{X}_2 + ... + W_L \overline{X}_L$$

أى يساوى :

$$\overline{X} = \sum_{h=1}^{L} W_h \overline{X}_h \qquad .... (5 - 11)$$

حيث

$$W_h = N_h / N$$

٥ - ٣ خطوات اختيار المعاينة الطبقية العشوائية :

يتطلب تصميم المعاينة الطبقية اتباع خطوات تصميم العينات بشكل عام مع ملاحظة وجود اختلافات فى طريقة اختيار الوحدات وتقدير معالم المجتمع ، أهمها .

- تقسيم المجتمع إلى طبقات بحيث تكون مفردات كل طبقة متجانسة فيما بينها لحد ما ، بينما نجد أن هناك فروقًا واضحة بين كل طبقة وأخرى .

- تقدير حجم العينة الطبقية الكلى للحصول على الدقة المطلوبة ، وهناك عدة صيغ لتحديد حجم العينة .
- توزيع حجم العينة على مختلف الطبقات بحيث تعطى أقل ما يمكن من أخطاء المعاينة لتكلفة ثابتة أو أقل تكلفة لتباين ثابت .
- يتم اختيار وحدات العينة من كل طبقة بشكل عشوائى (أى باستخدام إحدى طرق اختيار العينة العشوائية البسيطة أو بالأسلوب العشوائى المنتظم الذى سندرسه فيما بعد ، أو باستخدام الطرق الأخرى للاختيار العشوائى) .
- نقوم بتقدير أهم معالم المجتمع باستخدام بيانات جميع الوحدات المختارة من كل طبقة من طبقات المجتمع .

ويعد تقسيم المجتمع إلى طبقات من أهم الخطوات ، حيث يتوقف هذا التقسيم على درجة الدقة المطلوبة التى تعتمد على درجة التجانس داخل كل طبقة . ولا بد عند القيام بعملية تقسيم المجتمع إلى طبقات من الأخذ بالاعتبار إضافة لدرجة الدقة ، العوامل الأخرى كالإمكانات البشرية والفنية والمالية المخصصة للبحث .

أما حجم العينة الأمثل ، فهو الحجم الذى يعطينا أقصى دقة بأقل ما يمكن من التكاليف ، ولكن عملياً نجد أن حجم العينة الأمثل هو الذى يعطى أعلى دقة ممكنة بتكاليف محددة بصورة مسبقة .

ولتوزيع حجم العينة الإجمالى على مختلف الطبقات ، بحيث يعطى أقل ما يمكن من أخطاء المعاينة ، يوجد عدة طرق تسمى طرق تخصيص العينة وتتلخص فيما يلى :

طريقة التخصيص المتساوى :

يتم توزيع حجم العينة الإجمالى على مختلف الطبقات بشكل متساوٍ ، أى أن أحجام جميع الطبقات متساوية ، أى :

$n_1 = n_2 = n_3 = ... = n_L$

ويساوى حجم كل طبقة :

$$n_h = \frac{n}{L} \quad .... (5 - 12)$$

طريقة التخصيص المتناسب :

يتم توزيع حجم العينة الطبقية على مختلف الطبقات على أساس تناسب حجم الطبقة في المجتمع مع حجم المجتمع الإجمالي ، أي أن :

$$W_h = \frac{N_h}{N} = \frac{n_h}{n}$$

وبالتالي يكون حجم الطبقة (h) في العينة :

$$n_h = n\,\frac{N_h}{N} \qquad \text{.... (5 - 13)}$$

أي يساوي :

$$n_h = n\,W_h$$

طريقة التخصيص الأمثل :

يوزع حجم العينة الطبقية على الطبقات على أساس درجة تجانس هذه الطبقة وإدخال عامل التكاليف . فإذا كانت مفردات الطبقة متجانسة ، فإننا نختار عدداً أقل من الوحدات ، وكلما قل التجانس في مفردات الطبقة ازداد عدد الوحدات التي نختارها من الطبقة ، وذلك للتقليل من أخطاء المعاينة . ويمكننا القول إنه عند استخدام طريقة التوزيع الأمثل ، يكون حجم العينة من الطبقة كبيراً ، عندما يكون حجم الطبقة من المجتمع كبيراً أو تباين هذه الطبقة كبيراً ، أو يكونان كلاهما معاً كبيرين . وعند إدخال عامل التكاليف في تحديد حجم العينة في الطبقة ، نجد أن هذا الحجم يقل إذا كانت تكاليف الوحدة كبيرة ، والعكس بالعكس ، وذلك إضافة لحجم وتباين الطبقة في المجتمع .

ويتم بعد ذلك اختيار وحدات العينة من كل طبقة بالأسلوب العشوائي باستخدام إحدى طرق السحب العشوائي ، ثم نقوم بتقدير أهم معالم المجتمع ، ولا بد لنا من الإشارة إلى أن عدد العينات الممكنة يساوي حاصل ضرب $\binom{N_h}{n_h}$ لجميع الطبقات ، أي يساوي $\prod_{h=1}^{L}\binom{N_h}{n_h}$ حيث (π) ترمز إلى حاصل ضرب عدة أعداد .

٥ - ٤ تقدير معالم المجتمع باستخدام المعاينة الطبقية العشوائية :

٥ - ٤ - ١ تقدير متوسط المجتمع وتقدير القيمة الكلية للمجتمع :

إن الغاية الأساسية من استخدام أسلوب المعاينة ، تعميم نتائج العينة على المجتمع الذى اختيرت منه . ولنوضح الآن كيفية تقدير كلٌ من متوسط المجتمع والقيمة الكلية لمفردات المجتمع من بيانات العينة الطبقية .

إذا سحبنا عينة طبقية من مجتمع مكون من (L) طبقة ، يكون لدينا (L) متوسطًا للطبقات وهى : $\bar{x}_1 , \bar{x}_2 , \ldots , \bar{x}_L$ حيث :

متوسط الطبقة الأولى :

$$\bar{x}_1 = \frac{x_1}{n_1}$$

ومتوسط الطبقة الثانية :

$$\bar{x}_2 = \frac{x_2}{n_2}$$

ومتوسط الطبقة ذات الرتبة (h) :

$$\boxed{\bar{x}_h = \frac{x_h}{n_h}} \quad \ldots (5 - 14)$$

(حيث $\bar{x}_1 , \bar{x}_2 , \ldots , \bar{x}_h$ هى مجموع قيم العينة للطبقات 1,2,....,h على التوالى) .
إن متوسط العينة فى الطبقة (h) هو مقدر غير متحيز ومتسق لمتوسط الطبقة (h) فى المجتمع ، أى أن ($\bar{x}_h$) هو مقدر غير متحيز ومتسق لـ ($\bar{X}_h$) .

ويمكننا الحصول على تقدير القيمة الكلية للمجتمع ، وذلك بترجيح متوسطات الطبقات فى العينة بأحجامها فى المجتمع وذلك كما يلى :

– مقدر القيمة الكلية للطبقة الأولى فى المجتمع يساوى :

$$\hat{X}_1 = N_1 \bar{x}_1$$

وللطبقة الثانية :

$$\hat{X}_2 = N_2 \bar{x}_2$$

وبشكل عام للطبقة ذات الرتبة (h) :

$$\widehat{X}_h = N_h \, \bar{x}_h \qquad .... (5 - 15)$$

ويمكننا القول إن مقدر القيمة الكلية للمجتمع من عينة طبقية ولنرمز له بالرمز $(\widehat{X}_{st})$ يساوى مجموع تقديرات القيمة الكلية للطبقات ، أى يساوى :

$$\widehat{X}_{st} = \sum_{h=1}^{L} \widehat{X}_h$$

أى أن :

$$\widehat{X}_{st} = \sum_{h=1}^{L} N_h \, \bar{x}_h \qquad .... (5 - 16)$$

حيث $(h = 1, 2, ..., L)$.

٥ - ٤ - ٢ تقدير متوسط المجتمع على أساس عينة طبقية :

إذا رمزنا لمقدر متوسط المجتمع على أساس عينة طبقية بالرمز ($\bar{x}_{st}$) فإنه يساوى :

$$\bar{x}_{st} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2 + ... + N_L \bar{x}_L}{N_1 + N_2 + ... + N_L}$$

أى يساوى متوسطات الطبقات من العينة مرجحة بنسبة حجم الطبقة فى المجتمع إلى إجمالى حجم المجتمع ، أى :

$$\bar{x}_{st} = \frac{\sum_{h=1}^{L} N_h \, \bar{x}_h}{N} \qquad .... (5 - 17)$$

أى يساوى :

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \, \bar{x}_h \qquad .... (5 - 18)$$

حيث

$$W_h = \frac{N_h}{N} \quad \text{.... (5 - 1-9)}$$

ويعد متوسط العينة الطبقية ($\bar{x}_{st}$) مقدراً غير متحيز ومتسقًا لمتوسط المجتمع ، حيث نعلم أن توقع المقدر يجب أن يساوى متوسط المجتمع لكى يعد غير متحيز .

تطبيق (٥ – ١) :

يتكون مجتمع من الموظفين من (٦) موظفين يعملون فى الإدارتين (أ) و (ب) ، وكانت سنوات الخبرة لديهم :

$X_{11} = 2 \quad , X_{12} = 4 \quad , X_{13} = 6$
$X_{21} = 8 \quad , X_{22} = 12 \quad , X_{23} = 16$

المطلوب استخراج :

١ – الوسط الحسابى لسنوات الخبرة للموظف فى كل إدارة .
٢ – الوسط الحسابى لسنوات الخبرة للموظفين .
٣ – إجمالى عدد سنوات الخبرة لدى الموظفين .

الحـــل :

عدد سنوات الخبرة فى كلتا الإدارتين :
نستخدم الصيغة التالية :

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

ويكون عدد سنوات الخبرة فى الإدارة (أ) :

$$X_1 = X_{11} + X_{12} + X_{13}$$
$$= 2 + 4 + 6 = 12$$

وعدد سنوات الخبرة فى الإدارة (ب) :

$$X_2 = X_{21} + X_{22} + X_{23}$$
$$= 8 + 12 + 16 = 36$$

– إجمالى عدد سنوات الخبرة فى الإدارتين :

$$X = \sum_{h=1}^{L} X_h$$
$$= 12 + 36 = 48$$

الوسط الحسابى للطبقة (h) فى المجتمع :

$$\overline{X}_h = \frac{X_h}{N_h} = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h}$$

ويكون متوسط سنوات الخبرة فى الإدارة (أ) :

$$\overline{X}_1 = \frac{12}{3} = 4$$

ومتوسط سنوات الخبرة فى الإدارة (ب) :

$$\overline{X}_2 = \frac{36}{3} = 12$$

– متوسط المجتمع أى متوسط سنوات الخبرة للموظف سواء كان فى الإدارة (أ) أو الإدارة (ب) :

$$\overline{X} = \frac{X}{N} = \frac{\sum_{h=1}^{L} X_h}{N}$$

$$= \frac{12 + 36}{6} = \frac{48}{6} = 8$$

ونحصل على النتيجة نفسها باستخدام الصيغة التالية :

$$\overline{X} = \frac{\sum_{h=1}^{L} N_h \overline{X}_h}{N} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N}$$

$$= \frac{(3x4) + (3x12)}{6} = 8$$

أى (٨) سنوات .

تطبيق (٥ – ٢) :

سحبت عينة طبقية باستخدام بيانات المثال السابق (٥ – ١) حجمها $n = 4$ وزعت على الطبقات باستخدام التخصيص المتساوى $(n_1 = n_2 = 2)$. المطلوب :

– تحديد عدد العينات الممكن سحبها .

– استخراج متوسط مفردات العينات الممكن سحبها .

– تقدير متوسط المجتمع على أساس العينة الطبقية (استخدام بيانات العينة الطبقية الأولى الممكن سحبها) . ثم أثبت أن مقدر القيمة الكلية للمجتمع هو مقدر غير متحيز للقيمة الكلية للمجتمع (X) .

– توضيح علاقات كسر المعاينة فى العينة الطبقية المحسوبة وحساب $(\overline{x}_{st})$ على أساس عدم معرفة (N_1, N_2) .

الحــل :

– إن عدد العينات الممكن سحبها يساوى :

$$\prod_{h=1}^{L} \binom{N_h}{n}$$

أى يساوى :

$$= \binom{N_1}{n_1} \times \binom{N_2}{n_2}$$

$$= \binom{3}{2} \times \binom{3}{2} = 9$$

أى نستطيع اختيار إحدى العينات التسع الممكن سحبها ، حجم كل منها (٤) وحدات ، وحدتان منها من الطبقة الأولى ، ووحدتان من الطبقة الثانية .

- يتم عشوائياً اختيار الوحدات من كل طبقة من طبقات المجتمع .

- نستطيع أن نكون الجدول التالى الذى يوضح العينات التى يمكن سحبها وأهم البيانات والمقاييس المستخرجة منها .

No.	x_{h1}	x_{2i}	x_1	x_2	$\bar{x}_1$	$\bar{x}_2$	$N_1\bar{x}_1$	$N_2\bar{x}_2$	$\hat{X}_{st}$
1	2,4	8,12	6	20	3	10	9	30	39
2	2,4	8,16	6	24	3	12	9	36	45
3	2,4	12,16	6	28	3	14	9	42	51
4	2,6	8,12	8	20	4	10	12	30	42
5	2,6	8,16	8	24	4	12	12	36	48
6	2,6	12,16	8	28	4	14	12	42	54
7	4,6	8,12	10	20	5	10	15	30	45
8	4,6	8,16	10	24	5	12	15	36	51
9	4,6	12,16	10	28	5	14	15	42	57

من هذا الجدول نلاحظ ما يلى :

- تتألف العينة الطبقية الأولى من المفـردات (2,4, 8,12)

ومفردات العينة الطبقية الثانية الممكن سحبها (2,4, 8,16) .

وهكذا نجد أن كُلاً من العينات الممكن سحبها تتألف من أربع مفردات .

ونلاحظ فى العينة الأولى أننا سحبنا مفردتين من الطبقة الأولى (2,4) ومفردتين من الطبقة الثانية (8,12) وفى العينة الممكنة الثانية ، اخترنا مفردتين من الطبقة الأولى (2,4) ومفردتين من الطبقة الثانية (8,16) وهكذا لبقية العينات الممكنة . إن العينة الطبقية التى نختارها هى إحدى هذه العينات ، ولنفترض أن العينة الأولى التى وحداتها (2,4,8,12) هى العينة المختارة ، ولنقم بتقدير بعض معلمات المجتمع من بيانات هذه العينة .

لدينا

$$\bar{x}_h = \sum_{i=1}^{n_h} x_{hi}$$

ويمثل (x_h) القيمة الكلية للطبقة (h) من بيانات العينة ، فيكون لدينا القيم الكلية لبيانات الطبقة الأولى من العينة (x_1) ولبيانات الطبقة الثانية من العينة (x_2) حيث :

$$x_1 = x_{11} + x_{12}$$
$$= 2 + 4 = 6$$
$$x_2 = x_{21} + x_{22}$$
$$= 8 + 21 = 20$$

– لتقديـر متوسط قيـم كل طبقة من المجتمـع ، نستخـرج متوسط قيم كل طبقة من العينة $\bar{x}_h$ حيث :

$$\bar{x}_h = \frac{x_h}{n_h}$$

$$= \frac{\sum_{i=1}^{n_h} \bar{x}_{hi}}{n_h}$$

ومنه :

$$\bar{x}_1 = \frac{6}{2} = 3$$

$$\bar{x}_2 = \frac{20}{2} = 10$$

– إن مقدر متوسط قيم المجتمع من بيانات عينة طبقية يساوى :

$$\widehat{\bar{X}}_{st} = \bar{x}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{x}_h}{N}$$

وفى المثال نجد أن :

$$\bar{x}_{st} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N}$$

$$= \frac{\widehat{X}_{st}}{N}$$

ومن بيانات المثال نجد أن $N_1 = N_2 = 3$ وبالتالى نجد أن :

$$\bar{x}_{st} = \frac{(3x3)+(3x10)}{6}$$

$$= \frac{9+30}{6} = \frac{39}{6} = 6.5$$

– للبرهان على أن المقدر $\bar{x}_{st}$ هو تقدير غير متحيز لمتوسط المجتمع ، نعلم أن احتمال سحب أية عينة من العينات الممكنة يساوى :

$$\frac{1}{\binom{N_1}{n_1}\binom{N_2}{n_2} \cdots\cdots \binom{N_L}{n_L}}$$

وفى مثالنا يساوى هذا الاحتمال :

$$\frac{1}{\binom{N_1}{n_1}\binom{N_2}{n_2}} = \frac{1}{\binom{3}{2}\binom{3}{2}} = \frac{1}{9}$$

ونعلم أن المتوسط العام لقيم المجتمع :

$$\bar{X} = \frac{X}{N}$$

$$= \frac{2+4+6+8+12+16}{6}$$

$$= \frac{48}{6} = 8$$

ونريد أن نثبت أن :

$$E(\bar{x}_{st}) = \bar{X}$$

$$= 8$$

نعلم أن :

$$E(\bar{x}_{st}) = E\left(\frac{\hat{X}_{st}}{N}\right)$$

$$= \frac{1}{N} E(\hat{X}_{st})$$

$$= \frac{1}{N}\sum_{i=1}^{K} \hat{X}_i \, P(\hat{X}_i)$$

حيث K يرمز إلى عدد العينات الممكن سحبها (فى مثالنا K = 9) ، ومنه نجد أن :

$$E(\bar{x}_{st}) = \frac{1}{N}\left\{\left[\hat{X}_1 \; P(\hat{X}_1)\right] + \left[\hat{X}_2 \, P(\hat{X}_2)\right] + \ldots + \left[\hat{X}_K P(\hat{X}_K)\right]\right\}$$

$$E(\bar{x}_{st}) = \frac{1}{6}\left\{[39 + 45 + 51 + \ldots + 51 + 57] \times \frac{1}{9}\right\}$$

حيث $P(\hat{X}_K)$ متساوية لجميع العينات الممكنة وتساوى $\frac{1}{9}$

ويكون :

$$E(\bar{x}_{st}) = \frac{1}{6} \times 432 \times \frac{1}{9} = 8$$

وهى النتيجة نفسها التى توصلنا إليها عند حساب متوسط قيم المجتمع $(\bar{X})$ أى أن $(\bar{x}_{st})$ هو تقدير غير متحيز لمتوسط قيم المجتمع $(\bar{X})$.

كذلك نلاحظ أن القيمة الكلية للمجتمع تساوى (٤٨) ، ونجد أن :

$$E(\hat{x}_{st}) = (\hat{X}_1 + \hat{X}_2 + \ldots + \hat{X}_K)\, P(\hat{X})$$

$$= (39 + 45 + 51 + \ldots + 51 + 57) \times \frac{1}{9}$$

$$= 48$$

وهى النتيجة نفسها للقيمة الكلية للمجتمع (X) أى أن $(\hat{X}_{st})$ هو أيضاً تقدير غير متحيز لـ X .

فى حالة التوزيع (التخصيص) المتناسب ، نعلم أن كسر المعاينة لكل طبقة يساوى :

$$f_h = \frac{n_h}{n} = \frac{N_h}{N}$$

وعند استخراج متوسط العينة الطبقية $(\bar{x}_{st})$رجحنا متوسط الطبقة $(\bar{x}_h)$ بعدد مفردات المجتمع لكل طبقة من الطبقات أى بـ (N_h) وقسمنا الناتج على (N) . لذا يمكننا تقدير متوسط المجتمع على أساس بيانات العينة دون الحاجة إلى معرفة $N_1, N_2, \ldots$ ويساوى فى حالة التوزيع المتناسب :

$$\bar{x}_{st} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \ldots + n_L \bar{x}_L}{n}$$

أى أن :

$$\bar{x}_{st} = \frac{\sum_{h=1}^{L} n_h \bar{x}_h}{n} \quad .... (5 - 20)$$

وفى مثالنا نجد فى هذه الحالة أن :

$$\bar{x}_{st} = \frac{(2 \times 3) + (2 \times 10)}{4} = 6.5$$

وهو الجواب نفسه الذى حصلنا عليه سابقًا عند استخدام حجم الطبقات فى المجتمع .

إن هذا يعنى افتراضنا ثبات النسبة بين مفردات كل طبقة على أساس القيم $\frac{n_h}{n}$ وبين مفردات كل طبقة فى المجتمع $\frac{N_h}{N}$.

وباستخدام العلاقة التالية يمكننا استخراج كسر المعاينة كما يلى :

لدينا القيم $\frac{N_h}{N} = \frac{n_h}{n}$ ومنه نجد أن :

ومنه $n_h N = n N_h$:

$$\frac{n_h}{N_h} = \frac{n}{N} = f$$

أى أن :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = ... = \frac{N_L}{N_L} = \frac{n}{N}$$

وعند استخدام هذه الطريقة ، تسمى المعاينة الطبقية النسبية أو المعاينة الطبقية ذات كسر المعاينة المتساوى . وسنعود لشرح هذه الطريقة فى الصفحات القادمة .

٥ - ٤ - ٢ تباين التقديرات وتقديراتها :

أ - تباين الطبقة فى المجتمع :

لنرمز إلى التباين بين مفردات المجتمع داخل الطبقة (h) بالرمز (σ^2_h) حيث :

$$\sigma^2_h = \frac{1}{N_h} \sum_{i=1}^{N_h} (X_{hi} - \overline{X}_h)^2 \quad .... (5 - 21)$$

والتباين المعدل يساوى :

$$S^2_h = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \overline{X}_h)^2 \quad .... (5 - 22)$$

وعندما تكون (N_h) كبيرة فإن المقدار ($N_h - 1 \approx N_h$) وبالتالى نجد أن ($\sigma^2_h = S^2_h$) .

ب - تباين المجتمع :

إذا رمزنا إلى تباين المجتمع بالرمز (σ^2) نجد أن :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2 \quad .... (5 - 23)$$

حيث ($\overline{X}$) هو الوسط الحسابى للمجتمع . ويعنى ذلك أن تباين المجتمع (σ^2) يظهر تباين مفردات جميع الطبقات من المتوسط العام للمجتمع . ونعلم أن تباين الطبقة h (σ^2_h) يظهر تباين الطبقة من وسطها الحسابى ($\overline{X}_h$) . لذا يمكن القول إن تباين المجتمع يساوى مجموع تباينات الطبقات كلها محسوبة باستخدام متوسط المجتمع $\overline{X}$ أى أن :

$$\sigma^2 = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} (X_{hi} - \overline{X})^2 \quad .... (5 - 24)$$

كذلك نجد أن التباين المعدل يساوى :

$$S^2 = \frac{1}{N-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(X_{hi} - \overline{X})^2 \quad .... (5-25)$$

وفى المجتمعات الكبيرة نجد أن $(S^2 \simeq \sigma^2)$.

٤ - العلاقة بين تباين الطبقة وتباين المجتمع :

لتوضيح العلاقة بين تباين الطبقة فى المجتمع وتباين المجتمع (أى العلاقة بين σ^2_h و σ^2) ، نعلم أن :

$$\sigma^2 = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(X_{hi} - \overline{X})^2$$

وبإضافة وطرح $(\overline{X}_h)$ نجد أن :

$$\sigma^2 = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(X_{hi} - \overline{X}_h + \overline{X}_h - \overline{X})^2$$

$$= \frac{1}{N}\left[\sum_{h=1}^{L}\sum_{i=1}^{N_h}(X_{hi} - \overline{X}_h)^2 + \sum_{h=1}^{L}N_h(\overline{X}_h - \overline{X})^2\right]$$

$$= \frac{1}{N}\sum_{h=1}^{L}N_h\,\sigma_h^2 + \frac{1}{N}\sum_{h=1}^{L}N_h(\overline{X}_h - \overline{X})^2$$

وهكذا نلاحظ أن تباين المجتمع قد قسم إلى قسمين :

١ - التباين داخل الطبقة وهو عبارة عن الحد الأول ولنرمز له بالرمز (σ^2_w) .

٢ - التباين بين الطبقات ولنرمز له بالرمز (σ^2_b) أى أن :

$$\sigma^2 = \sigma_w^2 + \sigma_b^2 \quad .... (5-26)$$

حيث

$$\sigma_W^2 = \frac{1}{N} \sum_{h=1}^{L} N_h \sigma_h^2$$

$$\sigma_b^2 = \frac{1}{N} \sum_{h=1}^{L} N_h (\overline{X}_h - \overline{X})^2$$

وتوضح هذه العلاقة أنه كلما كان التباين فى الطبقة صغيراً فإن التباين بين الطبقات يكبر ، أى عندما تكون مفردات الطبقة متجانسة (أى لا يوجد فروق كبيرة بينها) فإن التباين داخل الطبقة (σ^2_w) يصغر وبالتالى يكبر التباين بين الطبقات (σ^2_b) . والعكس بالعكس ، إذا كانت مفردات كل طبقة غير متجانسة فإن التباين داخل الطبقة (σ^2_w) يكبر ويصغر التباين بين الطبقات (σ^2_b) .

تطبيق (٥ – ٣) :

لدينا مجتمع من الأشخاص مكون من (١٢) شخصاً مقسمين إلى (٣) طبقات حسب أعمارهم :

الطبقة الأولى :

$X_{11} = 6 , X_{12} = 10 , X_{13} = 2 , X_{14} = 4 , X_{15} = 8$

الطبقة الثانية :

$X_{21} = 9 , X_{22} = 18 , X_{23} = 12$

الطبقة الثالثة :

$X_{31} = 20 , X_{32} = 26 , X_{33} = 16 , X_{34} = 26$

المطلوب استخراج :

١ – تباين كل طبقة من الطبقات الثلاث .

٢ – تباين المجتمع (الكلى) .

٣ – توضيح العلاقة بين التباين داخل الطبقة والتباين بين الطبقات وتباين المجتمع الأصلى .

الحـل :

١ – تباين كل طبقة من الطبقات الثلاث للمجتمع :

نستخدم الصيغتين (21 - 5) و (22 -5) لاستخراج (σ^2 و S^2) . لذا لا بد من حساب متوسطات الطبقات فنجد أن :

$\overline{X}_1 = 6 , \overline{X}_2 = 13 , \overline{X}_3 = 22$

$N_1 = 5 , N_2 = 3 , N_3 = 4$

ويكون تباين الطبقات الثلاث كما يلى :

$$\sigma_1^2 = \frac{(6-6)^2 + (10-6)^2 + (2-6)^2 + (4-6)^2 + (8-6)^2}{5}$$

$$= \frac{0 + 16 + 16 + 4 + 4}{5} = \frac{40}{5} = 8$$

$$S_1^2 = \frac{40}{5-1} = 10$$

وبطريقة مماثلة نجد أن :

$$\sigma_2^2 = \frac{42}{3} = 14$$

$$S_2^2 = \frac{42}{2} = 21$$

$$\sigma_3^2 = \frac{72}{4} = 18$$

$$S_3^2 = \frac{72}{3} = 24$$

٢ – تباين المجتمع :

لا بد لنا من استخراج المتوسط العام للمجتمع والذى يساوى :

$$\overline{X} = \frac{\sum_{h=1}^{L} N_h \overline{X}_h}{N}$$

$$= \frac{(5 \times 6) + (3 \times 13) + (4 \times 22)}{12}$$

$$= \frac{157}{12} = 13.08$$

ويساوى تباين المجتمع :

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}$$

$$= \frac{(6 - 13.08)^2 + (10 - 13.08)^2 + ... + (26 - 13.08)^2}{12}$$

$$= \frac{722.91}{12} = 60.24$$

وبالتالى نجد أن :

$$S^2 = \frac{722.91}{12 - 1} = 65.72$$

- العلاقة بين التباين داخل الطبقات والتباين بين الطبقات :

نعلم أن :

$$\sigma^2 = \sigma_w^2 + \sigma_b^2$$

وأن :

$$\sigma_w^2 = \frac{1}{N} \sum_{h=1}^{L} N_h \, \sigma_h^2$$

$$= \frac{1}{12} \left[(5 \times 8) + (3 \times 14) + (4 \times 18)\right]$$

$$= \frac{1}{12} (40 + 42 + 72) = \frac{154}{12} = 12.83$$

و (σ^2_b) يساوى :

$$\sigma_b^2 = \frac{1}{N} \sum_{h=1}^{L} N_h \, (\overline{X}_h - \overline{X})^2$$

$$= \frac{1}{12} \quad \left[5 (6 - 13.08)^2 + 3 (13 - 13.08)^2 + 4 (22 - 13.08)^2\right]$$

$$= \frac{1}{12} (250.63 + 0.02 + 318.27)$$

$$= \frac{568.92}{12} = 47.41$$

ونجد أن :

$$\sigma^2 = \sigma_w^2 + \sigma_b^2$$

$$= 12.83 + 47.41 = 60.24$$

وهو الجواب السابق نفسه لـ σ^2 .

د - تباين تقدير متوسط المجتمع على أساس معاينة طبقية :

لنرمز إلى تباين تقدير متوسط المجتمع باستخدام بيانات عينة طبقية بالرمز $V(\bar{x}_{st})$ (أى مربع الخطأ المعيارى) ويساوى :

$$V(\bar{x}_{st}) = \frac{1}{K} \sum_{i=1}^{K} (\bar{x}_{st_i} - \bar{X})^2 \quad .... (5 - 27)$$

حيث (K) عدد العينات الممكنة و $\bar{x}_{st_i}$ هو متوسط العينة الممكنة سحبها (i) و $(\bar{X})$ المتوسط العام للمجتمع وعندما يكون عدد العينات الممكن سحبها كبيراً ، فإننا لا نستطيع حساب متوسطاتها ، لذا سنلجأ إلى حساب $V(\bar{x}_{st})$ بطريقة أخرى وذلك باستخدام تباين الطبقة (s_h^2) .

إذا رمزنا إلى $\frac{N_h}{N}$ بالرمز (W_h) يكون لدينا :

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$$

$$= W_1 \bar{x}_1 + W_2 \bar{x}_2 + ... + W_L \bar{x}_L$$

وبالتالى يكون تباين تقدير متوسط المجتمع :

$$V(\bar{x}_{st}) = W_1^2 V(\bar{x}_1) + W_2^2 V(\bar{x}_2) + ... + W_L^2 V(\bar{x}_L)$$

ونعلم أن تباين متوسط الطبقة (h) يساوى :

$$V(\bar{x}_h) = \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} \quad \text{.... (5 - 28)}$$

وتباين المتوسط العام يساوى :

$$V(\bar{x}) = \frac{N - n}{N} \frac{S^2}{n}$$

إذا اعتبرنا كل طبقة كمجتمع صغير وسحبنا عينة عشوائية بسيطة من كل طبقة فإن الصيغة (5 - 28) تستخدم لحساب تقدير تباين المتوسط فيكون :

$$V(\bar{x}_{st}) = W_1^2 \frac{N_1 - n_1}{N_1} \frac{S_1^2}{n_1} + W_2^2 \frac{N_2 - n_2}{N_2} \frac{S_2^2}{n_2}$$

$$+ \ldots\ldots + W_L^2 \frac{N_L - n_L}{N_L} \frac{S_L^2}{n_L}$$

$$= \sum_{h=1}^{L} \frac{N_h^2}{N^2} \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}$$

أى أن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} \quad \text{.... (5 - 29)}$$

كما أن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h S_h)^2}{n_h}$$

أى أن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{(N_h S_h)^2}{n_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2 \quad \text{.... (5 - 30)}$$

ويلاحظ من هذه الصيغة أن الحد الأول يظهر التباين عندما يكون السحب مع الإعادة أى أن معامل تصحيح المجتمع المحدود يساوى (١) . أما الحد الثانى فهو عبارة عن التصحيح الضرورى عندما يكون السحب بدون إعادة أو إذا كان المجتمع محدوداً .

تطبيق (٥ – ٤) :

مجتمع من الموظفين الموزعين إلى طبقتين حسب سنوات الخبرة :

الطبقة الأولى :

$$X_{11} = 1 , X_{12} = 3 , X_{13} = 5$$

الطبقة الثانية :

$$X_{21} = 10 , X_{22} = 16 , X_{23} = 22$$

سحبنا عينة حجمها (٤) موظفين موزعين بالتساوى على الطبقتين . المطلوب حساب تباين تقدير متوسط المجتمع والخطأ المعيارى للتقدير .

الحـل :

نعلم أن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{(N_h S_h)^2}{n_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

ونحتاج إلى حساب

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2$$

$$\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

ويكون :

$$\bar{X}_1 = \frac{1 + 3 + 5}{3} = 3$$

$$\bar{X}_2 = \frac{10 + 16 + 22}{3} = 16$$

$$S_1^2 = \frac{(1 - 3)^2 + (3 - 3)^2 + (5 - 3)^2}{3 - 1} = \frac{4 + 0 + 4}{2} = 4$$

$$S_2^2 = \frac{(10-16)^2 + (16-16)^2 + (22-16)^2}{3-1} = \frac{36+0+36}{2} = 36$$

ويكون $S_1 = 2, S_2 = 6$

ونجد أن $V(\bar{x}_{st})$ يساوى :

$$V(\bar{x}_{st}) = \frac{1}{6^2}\left[\frac{[(3 \times 2)^2}{2} + \frac{(3 \times 6)^2]}{2}\right] - \frac{1}{6^2}\left[[(3 \times 4) + (3 \times 36)]\right]$$

$$= \frac{1}{36}[18 + 162] - \frac{1}{36}[12 + 108]$$

$$= \frac{180}{36} - \frac{120}{36} = 5 - 3.333 = 1.667$$

ونلاحظ أنه عندما يكون السحب مع الإعادة (أى عندما يكون معامل تصحيح المجتمع المحدود يساوى "١") فإن تباين تقدير متوسط المجتمع يساوى خمسة ويكون المقدار (٣,٣٣٣) عبارة عن معامل التصحيح الضرورى عندما يكون السحب بدون إعادة أو إذا كان المجتمع محدوداً .

هـ - تباين تقدير متوسط المجتمع باستخدام بيانات عينة طبقية :

يتطلب حساب تباين تقدير متوسط المجتمع $V(\bar{x}_{st})$ باستخدام الصيغ السابقة حساب التباين المعدل لكل طبقة من طبقات المجتمع (S_h^2) وغالبًا ما يكون هذا التباين مجهولاً فى معظم التطبيقات العملية ، ولكننا نستطيع حساب تباين الطبقة من بيانات العينة الممثلة للمجتمع ، أى (s_h^2) حيث يعد هذا المقدر غير متحيز للتباين المعدل للمجتمع (S_h^2) :

$$s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(x_{hi} - \bar{x}_h)^2$$

ويكون تباين تقدير متوسط المجتمع المقدر من بيانات عينة طبقية ولنرمز له بالرمز $\hat{V}(\bar{x}_{st})$ مساويًا لـ :

$$\hat{V}(\bar{x}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h - n_h}{N_h}\frac{(N_h s_h)^2}{n_h} \quad \text{.... (5 - 31)}$$

وهو عبارة عن مقدر غير متحيز لـ $V(\bar{x}_{st})$

تطبيق (٥ – ٥) :

مجتمع مكون من (٦) أسر ، سحبت عينة حجمها (٤) أسر لتقدير متوسط الإنفاق الشهرى للأسر وكان الإنفاق الشهرى لأسر العينة كما يلى (بالاف الريالات) .

الطبقة الأولـى :

$$x_{11} = 2 , x_{12} = 4$$

الطبقة الثانية :

$$x_{21} = 8 , x_{22} = 16$$

المطلوب :

١ – استخراج تقدير متوسط المجتمع علماً بأن حجم المجتمع مقسوم بالتساوى بين الطبقتين .

٢ – استخراج $\hat{V}(\overline{x}_{st})$ أى التباين المقدر باستخدام بيانات العينة .

الحـل :

$$\hat{V}(\overline{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h \ s_h)^2}{n_h}$$

نقوم بحساب $\overline{x}_h$ و s_h^2 لكل من الطبقتين :

$$\overline{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (x_{hi} - \overline{x}_h)^2}{n_h - 1}$$

ومن بيانات المثال نجد أن :

$$\overline{x}_1 = \frac{2+4}{2} = 3 \quad \text{و} \quad \overline{x}_2 = \frac{8+16}{2} = 12$$

ويكون :

$$\bar{x}_{st} = \frac{(3 \times 3) + (12 \times 3)}{6} = 7.5$$

$$s_1^2 = \frac{1}{2-1}\left[(2-3)^2 + (4-3)^2\right]$$

$$= 1+1=2$$

$$s_2^2 = \frac{1}{2-1}\left[(8-12)^2 + (16-12)^2\right]$$

$$= 16 + 16 = 32$$

ويكون :

$$\hat{V}(\bar{x}_{st}) = \frac{1}{6^2}\left[\left(\frac{3-2}{3} \times \frac{3^2 \times 2}{2}\right) + \left(\frac{3-2}{3} \times \frac{3^2 \times 32}{2}\right)\right]$$

$$= \frac{1}{36}\left[\left(\frac{1}{3} \times \frac{18}{2}\right) + \left(\frac{1}{3} \times \frac{288}{2}\right)\right]$$

$$= \frac{1}{36}(3 + 48) = \frac{51}{36} = 1.417$$

و - تباين تقدير القيمة الكلية للمجتمع وتقديره باستخدام بيانات العينة :

نعلم أن تباين تقدير القيمة الكلية للمجتمع يساوى :

$$V(\hat{X}) = V(\hat{X}_{st}) = V(N\ \bar{x}_{st})$$

$$= N^2 V(\bar{x}_{st})$$

وباستخدام الصيغة السابقة المتعلقة بـ $V(\bar{x}_{st})$ نجد أن :

$$V(\hat{X}_{st}) = N^2\left[\frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h - n_h}{N_h}\frac{(N_h S_h)^2}{n_h}\right]$$

ومنه نجد أن تباين تقدير القيمة الكلية للمجتمع يساوى :

$$V(\widehat{X}_{st}) = \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h S_h)^2}{n_h} \quad .... (5 - 32)$$

وفى حالة عدم معرفة S_h^2 نستخدم تباين العينة ، ويكون تقدير تباين تقدير القيمة الكلية للمجتمع :

$$\widehat{V}(\widehat{X}_{st}) = \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h s_h)^2}{n_h} \quad .... (5 - 33)$$

حيث $\widehat{V}(\widehat{X}_{st})$ هو مقدر غير متحيز لـ $V(\widehat{X}_{st})$ ، و (s_h^2) هو تباين الطبقة (h) من بيانات العينة أى :

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$$

تطبيق (٥ – ٦) :

على ضوء بيانات المثال السابق ، المطلوب استخراج تباين تقدير القيمة الكلية للمجتمع المقدر من بيانات العينة .

الحـل :

باستخدام الصيغة (33 - 5) نجد أن :

$$\hat{V}(\widehat{X}_{st}) = \left[\frac{(3 - 2)}{3} \frac{(3^2 \times 2)}{2} + \frac{(3 - 2)}{3} \frac{(3^2 \times 32)}{2} \right]$$

$$= 3 + 48 = 51$$

ويكون تقدير الانحراف المعيارى للقيمة الكلية المقدرة :

$$\sqrt{\widehat{V}(\widehat{X}_{st})} = \sqrt{51} = 7.14$$

٥ - ٥ حدود الثقة لتقدير متوسط المجتمع وتقدير القيمة الكلية للمجتمع:

عند استخدام بيانات العينة الطبقية لتقدير كل من متوسط المجتمع والقيمة الكلية للمجتمع ، نستطيع استخراج حدود الثقة بمستوى ثقة محدد % (1-α) وذلك باستخدام الصيغ التالية :

– حدا الثقة لتقدير متوسط المجتمع بمستوى ثقة % (1-α) :

$$\bar{x}_{st} + Z_{(\alpha/2)} \sqrt{\hat{V}\,(\bar{x}_{st})} \leq \mu < \bar{x}_{st} + Z_{(1-\alpha/2)} \sqrt{\hat{V}\,(\bar{x}_{st})} \quad \text{.... (5 - 34)}$$

إذا كان حجم العينة كبيراً (٣٠ فاكثر) . ونستخدم الصيغة :

$$\bar{x}_{st} + t_{(\alpha/2,\, n-1)} \sqrt{\hat{V}\,(\bar{x}_{st})} \leq \mu < \bar{x}_{st} + t_{(1-\alpha/2,\, n-1)} \sqrt{\hat{V}\,(\bar{x}_{st})} \quad \text{.... (5 - 35)}$$

إذا كان حجم العينة أقل (٣٠) وحدة حيث :

$$\hat{V}\,(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h\, s_h)^2}{n_h}$$

– حدا الثقة لتقدير القيمة الكلية للمجتمع بمستوى ثقة % (1-α) :

$$\hat{X}_{st} \mp Z_{1-\alpha/2} \sqrt{\hat{V}\,(N\,\bar{x}_{st})} \quad \text{.... (5 - 36)}$$

حيث

$$\hat{V}\,(N\,\bar{x}_{st}) = N^2\ \hat{V}\,(\bar{x}_{st})$$

$$\hat{X}_{st} \mp t_{(\alpha/2,\, n-1)} \sqrt{\hat{V}\,(\bar{x}_{st})} \quad \text{.... (5 - 37)}$$

وتستخرج قيمة (Z) فى جميع الحالات من جداول توزيع المنحنى الطبيعى وقيمة (t) من جداول توزيع ستيودنت حيث :

$$Z_{\alpha/2} = -Z_{1-\alpha/2}$$

$$t_{\alpha/2,\, n-1} = -t_{1-\alpha/2,\, n-1}$$

تطبيق (٥ – ٧) :

تتكون إحدى المدن من (٣١٠٠) أسرة اختيرت منها عينة حجمها (٤٠٠) أسرة لتقدير متوسط الإنفاق الشهرى للأسرة . إذا كانت الأسر مقسمة إلى ثلاث طبقات حسب مستوى الدخل وكانت لدينا البيانات التالية :

	الطبقة (١)	الطبقة (٢)	الطبقة (٣)	الإجمالى
حجم المجتمع	١٥٥٠	٦٢٠	٩٣٠	٣١٠٠
حجم العينة	٢٠٠	٨٠	١٢٠	٤٠٠
متوسط الإنفاق (بالريالات)	٤٠٠٠	٨٠٠٠	١٥٠٠٠	
تباين العينة (s^2)	٣٦٠٠	٦٤٠٠	١٢١٠٠	

المطلوب :

– تقدير متوسط الإنفاق الشهرى فى هذه المدينة بمستوى ثقة (٩٥٪) .

– تقدير إجمالى إنفاق المدينة بمستوى ثقة (٩٥٪) .

الحـل :

$$\bar{x}_{st} = \frac{1}{N} \sum_{h=1}^{L} (N_h \bar{x}_h)$$

$$= \frac{1}{N} (N_1 \bar{x}_1 + N_2 \bar{x}_2 + N_3 \bar{x}_3)$$

$$= \frac{1}{3100} [(1550 \times 4000) + (620 \times 8000) + (930 \times 15000)]$$

$$= \frac{25110000}{3100} = 8100$$

ولتقدير متوسط الإنفاق بمستوى ثقة (٩٥٪) نوجد حدى الثقة .

$$\bar{x}_{st} + Z_{(\alpha/2)} \sqrt{\hat{V}(\bar{x}_{st})} \leq \mu < \bar{x}_{st} + Z_{(1-\alpha/2)} \sqrt{\hat{V}(\bar{x}_{st})}$$

$$\hat{V}(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{(N_h\ s_h)^2}{n_h}$$

$$= \frac{1}{(3100)^2} \left[\left(\frac{1550 - 200}{1550} \times \frac{(1550)^2 \times 3600}{200}\right) + \left(\frac{620 - 80}{620} \times \frac{(620)^2 \times 6400}{80}\right)\right.$$

$$\left. + \left(\frac{930 - 120}{930} \times \frac{(930)^2 \times 12100}{120}\right)\right] = \frac{1}{9610000} [37665000 + 26784000 + 75957750]$$

$$= \frac{140406750}{9610000} = 14.61$$

ويكون

$$\sqrt{\hat{V}(\bar{x}_{st})} = \sqrt{14.61} = 3.82$$

وحيث

$$- Z_{0.975} = Z_{0.025} = - 1.96$$

يكون حدا الثقة

$$8100 - 1.96 \times 3.82 \leq \mu \leq 8100 + 1.96 \times 3.82$$

$$8100 - 7.49 \leq \mu \leq 8100 + 7.49$$

$$8092.5 \leq \mu \leq 8107.49$$

أى بمستوى ثقة (٩٥٪) فإن متوسط إنفاق المدينة سيقع بين (8092.51) و (8107.49) ريالات ويمكننا القول إنه لو سحبنا عدداً كبيراً من العينات من الأسر حجم كل منها (٤٠٠) أسرة ، وحسبنا حدود الثقة لكل عينة ، فإن (٩٥٪) من هذه الحدود ستتضمن متوسط إنفاق الأسرة فى المدينة .

- أما تقدير إجمالى إنفاق المدينة فيكون :

$$\hat{X}_{st} + Z_{\alpha/2}\sqrt{\hat{V}\ (N\bar{x}_{st})} \le X \le \hat{X}_{st} + Z_{1-\alpha/2}\sqrt{\hat{V}\ (N\bar{x}_{st})}$$

$$\hat{X}_{st} = N\bar{x}_{st} = 3100 \times 8100 = 25110000$$

$$\hat{V}\ (N\bar{x}_{st}) = N^2\,\hat{V}\ (\bar{x}_{st}) = (3100)^2\ (14.61)^2 = 2051274681$$

ويكون

$$\sqrt{\hat{V}\ (N\bar{x}_{st})} = 45291$$

وبالتبديل نجد أن :

$$25110000 - 1.96 \times 45291 \le X \le 25110000 + 1.96 \times 45291$$

$$25110000 - 88770 \le X \le 25110000 + 88770$$

$$25021230 \le X \le 25198770$$

أى أن إجمالى إنفاق المدينة بمستوى ثقة (٩٥٪) يتراوح بين (٢٥٠٢١٢٣٠) ريالاً و (٢٥١٩٨٧٧٠) ريالاً .

٥ - ٦ طرق تخصيص حجم العينة على الطبقات وتحديد حجم العينة :

تتركز المشكلة الأساسية التى تواجه مصمم البحث فى تحديد حجم العينة المناسب وتخصيص حجم كل طبقة من الطبقات ، إذ نجد أن حجم العينة الإجمالى وحجم كل طبقة من طبقات العينة يؤثران على تقديرات متوسط المجتمع والتباين . وقد ذكرنا سابقًا بأن تحديد حجم العينة يتم على أساس الحصول على أقصى دقة ممكنة بأقل ما يمكن من التكاليف ، وفى الحياة العملية كثيراً ما تحدد التكاليف المخصصة للبحث أولاً ، ومن ثم يحدد حجم العينة الذى يحقق أعلى دقة ممكنة وفق الإمكانات المالية المخصصة .

ولابد لنا من التنويه بأن تكاليف المعاينة هى عبارة عن تكاليف تصميم العينة ، وتكاليف تجهيز إطار البحث ، وتدريب الباحثين ، بالإضافة إلى نفقات جمع وتبويب البيانات التى تم الحصول عليها ، والنفقات الإدارية ، والنفقات الأخرى . وبشكل عام ، يمكننا تقسيم نفقات المعاينة إلى قسمين رئيسيين :

- نفقات ثابتة لا تتأثر بحجم العينة ولنرمز لها بالرمز (C_0) مثل نفقات تصميم البحث خاصة الإدارية منها .

- نفقات غير ثابتة تتوقف على حجم العينة فى كل طبقة مثل نفقات جمع البيانات وطباعة الاستمارات وغيرها . إذا رمزنا إلى ما تتطلبه كل وحدة من نفقات فى الطبقة (h) بالرمز (C_h) ، نستطيع صياغة دالة تكاليف المعاينة (Sampling Cost Function) بالشكل :

$$C = C_o + \sum_{h=1}^{L} n_h C_h \qquad \text{.... (5 - 38)}$$

حيث (C) تمثل إجمالى تكاليف المعاينة . ويمكننا صياغة هذه الدالة بأشكال أخرى حسب تكلفة الوحدة فى الطبقة ، مثلاً قد تكون هذه التكلفة متساوية لجميع وحدات الطبقات ولا يوجد فروق بينها لذا نستخدم صيغة أخرى للدالة تختلف عن الصيغة السابقة .

وبشكل عام فإن الصيغة المستخدمة لتحديد حجم العينة الإجمالى (حسب طريقة التخصيص المتناسب أو المتساوى) لتقدير متوسط المجتمع (μ) أو لتقدير القيمة الكلية (X) إذا كان خطأ التقدير المطلوب (β) هى :

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 S_h^2}{W_h}}{N^2 D + \sum_{h=1}^{L} N_h S_h^2} \qquad \text{.... (5 - 39)}$$

حيث (W_h) هو كسر يمثل نسبة عدد مشاهدات الطبقة (h) إلى الإجمالى $\left(W_h = \frac{N_h}{N}\right)$ وأن (S_h^2) هو تباين المجتمع المعدل للطبقة (h) . كما أن :

$D = \frac{\beta^2}{Z^2}$ عندما نريد تقدير (μ) .

$D = \frac{\beta^2}{Z^2 N^2}$ عندما نريد تقديرالقيمة الكلية (X) .

ويمكننا استخدام تباين العينة للطبقة h (s_h) كمقدر لـ (S_h) .

(يمكن الرجوع إلى الملحق رقم (3 - 5) لتوضيح كيفية الحصول على الصيغة المستخدمة لتحديد حجم العينة . وكثير من الإحصائيين يضعون (Z = 2) عند مستوى ثقة

(٩٥٪) تســاوى $(Z = 1.96 \simeq 2)$ أى تقريباً (٢) وبالتالى تصبح $\left(D = \frac{B^2}{4} \text{ or } D = \frac{B^2}{4N^2} \right)$

وبعد أن يتم تحديد حجم العينة ، يتم تخصيص حجم العينة من كل طبقة من الطبقات وذلك باستخدام إحدى الطرق التالية :

١ – التخصيص المتناسب .

٢ – التخصيص المتساوى .

٣ – التخصيص الأمثل .

٤ – تخصيص نيمان .

وسنقوم باستخراج الصيغة المناسبة لتحديد حجم العينة وتوزيعها على الطبقات حسب طريقة التخصيص المستخدمة .

٥ – ٦ – ١ طريقة التخصيص المتناسب : Proportional Allocation Method

تعد طريقة التخصيص المتناسب لتحديد حجم العينة من كل طبقة من الطبقات ، من الطرق الشائعة الاستخدام ، نظراً لعدم إدخال عامل التكاليف فى الصيغ المتعلقة بهذه الطريقة مما يؤدى إلى سهولتها إذا قورنت بالصيغ الأخرى .

يتم تقسيم حجم العينة الإجمالى على مختلف الطبقات على أساس نسبة ثابتة هى كسر المعاينة :

$$f = \frac{n}{N} = \frac{n_h}{N_h}$$

أى أن :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots\cdots = \frac{n_h}{N_h} = \frac{n}{N} = f$$

أى أن :

$$n_h = n\,\frac{N_h}{N}$$

ويمكننا القول إن كل طبقة تشكل مجتمعاً صغيراً يتم اختيار عينة من وحداته بشكل عشوائى . وبذلك يكون احتمال سحب وحدة معاينة من الطبقة (h) يساوى كسر المعاينة أى :

$$P(X_{hi}) = f = \frac{n_h}{N_h} = \frac{n}{N}$$

وهذا الاحتمال متشابه فى جميع وحدات الطبقة الواحدة ، وتستخدم الصيغة نفسها لحساب احتمال سحب وحدات المعاينة فى جميع الطبقات .

تقديرات التخصيص المتناسب :

ذكرنا أن مقـدر متوسط المجتمع على أساس معاينة طبقية يساوى :

$$\bar{x}_{st} = \frac{\sum_{h=1}^{L} N_h \, \bar{x}_h}{N_h}$$

وحسب طريقة التخصيص المتناسب ، نعلم أن $N_h = \frac{n_h}{f}$ ، ونجد أن مقدر متوسط المجتمع باستخدام طريقة التخصيص المتناسب ، ولنرمز له بالرمز ($\bar{x}_{prop}$) يساوى :

$$\bar{x}_{prop} = \frac{\sum_{h=1}^{L} \frac{n_h \, \bar{x}_h}{f}}{\sum_{h=1}^{L} \frac{n_h}{f}}$$

$$= \frac{\sum_{h=1}^{L} n_h \, x_h}{n}$$

ومنه نجد أن :

$$\boxed{\bar{x}_{prop} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{n_h} x_{hi}}{n}} \qquad \text{.... (5 - 40)}$$

أى أن مقدر متوسط المجتمع على أساس معاينة طبقية حسب طريقة التخصيص المتناسب هو عبارة عن الوسط الحسابى للعينة (غير المرجح) .

ولحساب تباين تقدير متوسط المجتمع باستخدام طريقة التخصيص المتناسب ، نعلم أنه إذا كان معامل تصحيح المجتمع المحدود لا يساوى الواحد فإن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 \; S_h^2}{n_h}$$

ويساوى هذا التباين إذا كان معامل تصحيح المجتمع المحدود مساوياً الواحد :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h}$$

وحسب طريقة التخصيص المتناسب ، نعلم أن $(n_h = n \frac{N_h}{N})$ وبتبديل قيمة (n_h) فى الصيغ السابقة نجد أن تبـاين تقديـر متوسط المجتمـع للتخصيص المتنـاسب ولنرمـز لـه بـ $V(\bar{x}_{prop})$ يساوى :

$$V(\bar{x}_{prop}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - \frac{n_h}{N} n}{N_h} \frac{N_h^2 S_h^2}{\frac{N_h n}{N}}$$

وبوضع (N_h) خارج قوس ، وبتبسيط الصيغة السابقة نجد أن :

$$V(\bar{x}_{prop}) = \frac{N - n}{N} \sum_{h=1}^{L} \frac{N_h}{N} \frac{S_h^2}{n} \quad .... (5 - 41)$$

وعندما يكون معامل تصحيح المجتمع المحدود مساوياً للواحد ، نجد أن :

$$V(\bar{x}_{prop}) = \sum_{h=1}^{L} \frac{N_h}{N} \frac{S_h^2}{n} \quad .... (5 - 42)$$

وعندما يكون تباين المجتمع المعدل للطبقة مجهولاً ، نضع تقديره (s_h^2) ويصبح تقدير تباين متوسط المجتمع $\hat{V}(\bar{x}_{prop})$ حيث نضع (s_h^2) عوضاً عن (S_h^2) فى الصيغتين السابقتين .

ولتحديد حجم العينة حسب طريقة التخصيص المتناسب نعلم أن :

$$D = \frac{B^2}{Z^2} = Var(\bar{x}_{prop})$$

أى أن :

$$D = \frac{N - n}{N} \sum_{h=1}^{L} \frac{N_h}{N} \frac{S_h^2}{n}$$

ومنه نجد بعد إجراء بعض العمليات الرياضية أن حجم العينة حسب طريقة التخصيص المتناسب يساوى :

$$n = \frac{N \sum_{h=1}^{L} N_h S_h^2}{N^2 D + \sum_{h=1}^{L} N_h S_h^2} \quad \text{.... (5 - 43) .}$$

ويمكننا استخدام (n_o) كتقريب أو لتحديد حجم العينة حيث

$$n_o = \frac{\sum_{h=1}^{L} W_h S_h^2}{D}$$

وإذا كان كسر المعاينة (n_o/N) أكبر من (٥٪) (أو ١٠٪ أحياناً) نستخرج حجم العينة النهائى .

$$n = \frac{n_o}{1 + (n_o / N)}$$

ويتم توزيعه على الطبقات باستخدام الصيغة ($n_h = n \frac{N_h}{N}$) .

(تطبيق ٥ – ٨) :

تمثل البيانات التالية عدد أفراد أسر (١٠) موظفين موزعين إلى طبقتين :

الطبقة الأولى : ٢ ، ٤ ، ٢ ، ٣ ، ٦ ، ٧ .

الطبقة الثانية : ٦ ، ٨ ، ٢ ، ٢ .

وقد تم اختيار عينة مفرداتها (٣ ، ٦ ، ٢) من الطبقة الأولى و(٦ ، ٨) من الطبقة الثانية .

المطلوب استخراج :

١ – تقدير متوسط المجتمع (باستخدام طريقة التخصيص المتناسب) .

٢ – تباين تقدير متوسط المجتمع .

٣ – تقدير تباين تقدير متوسط المجتمع .

الحـــل :

١ – يلاحظ أن طريقة التخصيص المتناسب هى الطريقة المستخدمة لتخصيص العينة على الطبقات : نعلم أن :

$$f = \frac{n}{N} = \frac{n_h}{N_h} = \frac{n_1}{N_1} = \frac{n_2}{N_2}$$

$$= \frac{5}{10} = \frac{3}{6} = 2/4 = 1/2$$

ويكون

$$\overline{x}_{prop} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{n_h} x_{hi}}{n}$$

$$= \frac{1}{5}(3 + 6 + 2 + 6 + 8) = \frac{25}{5} = 5$$

وهو تقدير متوسط المجتمع .

أما متوسط المجتمع فيساوى :

$$\overline{X} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{n_h} X_{hi}$$

$$= \frac{1}{10}\ (2 + 4 + ... + 3 + 3)$$

$$= \frac{44}{10}\ = 4.4$$

ويلاحظ أن احتمال سحب أية وحدة معاينة يساوى (f) أى (1/2) .

٢ – لاستخراج تباين تقدير متوسط المجتمع فى حالة التخصيص المتناسب نستخدم الصيغة :

$$V(\overline{x}_{prop}) = \frac{N - n}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} \frac{S_h^2}{n}$$

إذا كان معامل تصحيح المجتمع المحدود لا يساوى الواحد ، ونهمل هذا المعامل إذا كان مساويًا للواحد . لذا نحتاج إلى حساب تباين كل طبقة , S^2 أى S^2_2 و S^2_1 وذلك باستخدام الصيغة التالية :

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \overline{X}_h)^2$$

فنجد أن $S_1^2 = 4.4 \quad , S_2^2 = 6$

ويكون التباين لتقدير متوسط المجتمع للتخصيص المتناسب :

$$V(\overline{x}_{prop}) = \frac{10-5}{10}\left[\left(\frac{6}{10} \times \frac{4.4}{5}\right) + \left(\frac{4}{10} \times \frac{6}{5}\right)\right]$$

$$= \frac{5}{10}\left[0.528 + 0.\ 48\right]$$

$$= \frac{5}{10}\left[1.008\right] = 0.504$$

ويساوى هذا التباين (1.008) إذا كان معامل تصحيح المجتمع المحدود مساويًا للواحد .

٢ - إذا كانت مفردات المجتمع غير معلومة ، نستخدم مفردات العينة ويكون تقدير تباين متوسط المجتمع للتخصيص المتناسب $\hat{V}(\overline{x}_{prop})$ ، ونحسب تباين الطبقات من بيانات العينة .

$$s_1^2 = 4.33 \quad , s_2^2 = 2$$

ويكون تقدير التباين :

$$\hat{V}(\overline{x}_{prop}) = \frac{10-5}{10}\left[\left(\frac{6}{10} \times \frac{4.33}{5}\right) + \left(\frac{4}{10} \times \frac{2}{5}\right)\right]$$

$$= \frac{5}{10}\left[0.52 + 0.16\right]$$

$$= \frac{5}{10}\ (0.68) = 0.34$$

ويساوى تقدير هذا التباين (٠،٦٨) عند إهمال معامل تصحيح المجتمع المحدود .

٥ - ٦ - ٢ طريقة التخصيص المتساوى : (Equal Allocation Method)

إن حجم الطبقة (h) فى العينة حسب طريقة التخصيص المتساوى هو :

$$n_h = \frac{n}{L}$$

ونعلم أن تباين متوسط العينة الطبقية يساوى :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 S_h^2}{n_h}$$

وبتعويض (n_h) بقيمتها حسب هذه الطريقة من التخصيص ، نجد أن تباين متوسط العينة الطبقية وفقًا لطريقة التخصيص المتساوى ($V(\bar{x}_{eq})$ يساوى :

$$V(\bar{x}_{eq}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n/L} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2 \qquad \text{.... (5 - 44)}$$

ونريد تحديد حجم العينة (n) بحيث يكون التباين السابق يساوى D أى أن :

$$D = \frac{B^2}{Z^2} = V(\bar{x}_{eq})$$

نعلم من الصيغة ($V(\bar{x}_{eq}$ وباستخدام (D) أن :

$$N^2 D = \frac{L}{n} \sum_{h=1}^{L} N_h^2 S_h^2 - \sum_{h=1}^{L} N_h S_h^2$$

ومنه نجد أن :

$$N^2 D + \sum_{h=1}^{L} N_h S_h^2 = \frac{L}{n} \sum_{h=1}^{L} N_h^2 S_h^2$$

ونجد أن حجم العينة حسب طريقة التخصيص المتساوى :

$$n = \frac{L \sum_{h=1}^{L} N_h^2 S_h^2}{N^2 D + \sum_{h=1}^{L} N_h S_h^2} \qquad \text{.... (5 - 45)}$$

حيث $D = (B)^2 = \frac{B^2}{Z^2}$ هو خطأ التقدير المطلوب و Z القيمة المقابلة لمستوى ثقة معين في جداول التوزيع الطبيعي) .

وعندما يكون معامل تصحيح المجتمع المحدود مساوياً للواحد فإن صيغة تباين المتوسط تصبح :

$$V(\bar{x}_{eq}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h}$$

وبالتالي يصبح حجم العينة بعد تبديل (n_h) بقيمتها :

$$n = \frac{L \sum_{h=1}^{L} N_h^2 S_h^2}{N^2 D} \quad .... (5 - 46)$$

وعندما يكون التباين (S_h^2) مجهولاً ، نضع تقديره من عينة طبقية أي نضع (s_h^2) .

تطبيق (٥ – ٩) :

نريد اختيار عينة طبقية لتقدير سنوات الخبرة للموظفين في إحدى الجهات التي يبلغ عدد موظفيها (١٠٠) موظف موزعين بالتساوي إلى طبقتين هما : الموظفون الإداريون والموظفون الفنيون . إذا كان لدينا البيانات التالية من دراسة سابقة :

$$\bar{x} = 7.28 \qquad V(\bar{x}_{eq}) = 0.28$$

$$S_1^2 = 2 \qquad S_2^2 = 4.5$$

ما هو حجم العينة المناسب وحجم كل طبقة حسب طريقة التخصيص المتساوي ، علماً بأن خطأ التقدير المطلوب هو (1.5) .

باستخدام الصيغة رقم (45 - 5) نجد أن :

$$n = \frac{2\left[(50^2 \times 2) + (50^2 \times 4.5)\right]}{\frac{\left[(100)^2 (1.5)^2\right]}{4} + \left[(50 \times 2) + (50 \times 4.5)\right]} = 5.46$$

أي أن حجم العينة تقريباً هو (٦) موظفين ويكون :

$$n_1 = n_2 = \frac{6}{2} = 3$$

٥ - ٦ - ٢ طريقة التخصيص الأمثل : (Optimum Allocation)

تعتمد هذه الطريقة على إدخال عامل التكاليف والدقة ، وذلك عند تخصيص حجم كل طبقة ، لاختلاف هذه التكاليف من طبقة لأخرى فى بعض الأحيان . مثلا نجد أن تكاليف جمع البيانات من وحدات تقع فى المناطق النائية تتطلب نفقات إضافية قد تبلغ أضعاف ما تتكلفه هذه العملية فى المدن بسبب ارتفاع نفقات السفر وغيرها .

ويمكننا القول إننا نريد تحديد حجم (n_h) بشكل يكون فيه $V(\overline{x}_{st})$ أقل ما يمكن ، باستخدام نفقات محددة ، كما يمكن أيضًا تحديد حجم الطبقة (n_h) بحيث تكون التكلفة أقل ما يمكن لتباين محدد . ويتم تحديد حجم الطبقة (h) أى (n_h) بالعلاقة التالية وذلك باستخدام دالة لاغرانج (Lagrange) فى صيغة $V(\overline{x}_{st})$ لإيجاد أقل قيمة ممكنة للتكاليف * .

$$n_h = n \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} (N_h S_h / \sqrt{C_h})} \qquad .... (5 - 47)$$

حيث (c_h) تمثل نفقات الوحدة فى الطبقة (h) :

يتضح من هذه الصيغة أننا نأخذ من طبقة ما عينة حجمها كبير إذا كان حجم الطبقة فى المجتمع كبيراً أو إذا كان التباين بين وحدات هذه الطبقة كبيراً أو إذا كانت تكاليف هذه الطبقة قليلة أو إذا تحققت جميع هذه العوامل مع بعضها . أى أنه كلما كان حجم الطبقة (N_h) كبيراً وكلما كانت قيم الطبقة غير متجانسة ، يجب أن يكون حجم العينة كبيراً . كذلك يجب أن يكون حجم العينة كبيراً عندما تكون نفقات وحدة المعاينة لكل وحدة صغيرة والعكس بالعكس .

تقديرات التخصيص الأمثل :

أ - تقدير متوسط المجتمع :

نستخدم الصيغة التالية لتقدير متوسط المجتمع حسب طريقة التخصيص الأمثل ولنرمز له بالرمز $(\overline{x}_{opt})$.

$$(\overline{x}_{opt}) = \frac{\sum_{h=1}^{L} N_h \overline{x}_h}{N} \qquad .... (5 - 48)$$

* انظر الملحق رقم (٥-٢) .

أى هى الصيغة نفسها المستخدمة عند حساب متوسط عينة طبقية ($\bar{x}_{st}$) .

ب - تباين تقدير متوسط المجتمع وتقديره :

لنرمز إلى تباين تقدير متوسط المجتمع (مربع الخطأ المعيارى) المحسوب على أساس التخصيص الأمثل بالرمز ($\bar{x}_{opt}$) V ونستخرج صيغته كما يلى :

لدينا

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - S_h}{N_h} \frac{N_h^2 S_h^2}{n_h}$$

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{n_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

وبتبديل قيمة (n_h) بما تساويه من الصيغة (47 - 5) نجد أن تباين تقدير المتوسط يساوى :

$$V(\bar{x}_{opt}) = \frac{1}{N^2} \frac{1}{n} \sum_{h=1}^{L} N_h^2 S_h^2 \frac{\sum_{h=1}^{L} (N_h S_h)/\sqrt{C_h}}{N_h S_h/\sqrt{C_h}} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

أى أنه :

$$V(\bar{x}_{opt}) = \frac{1}{N^2} \frac{1}{n} \left(\sum_{h=1}^{L} N_h S_h \sqrt{C_h} \right) \left(\sum_{h=1}^{L} \frac{(N_h S_h)}{\sqrt{C_h}} \right) - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2 \qquad \text{.... (5 - 49)}$$

أما تقدير تباين متوسط المجتمع المقدر حسب التخصيص الأمثل ، فيتم حسابه باستخدام العلاقة السابقة نفسها ، مع تبديل (S_h^2) بـ (s_h^2) أى تباين الطبقة (h) من العينة

$$\hat{V}(\bar{x}_{opt}) = \frac{1}{N^2} \frac{1}{n} \left(\sum_{h=1}^{L} N_h s_h \sqrt{C_h} \right) \left(\sum_{h=1}^{L} \frac{N_h s_h}{\sqrt{C_h}} \right) - \frac{1}{N^2} \sum_{h=1}^{L} N_h s_h^2 \qquad \text{.... (5 - 50)}$$

حيث :

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)$$

جـ - تحديد حجم العينة :

لتحديد حجم العينة نعلم أن : $D = \frac{B^2}{Z^2} = V(\bar{x}_{opt})$ ، ونعوض فى الصيغة (49 - 5)

ونبسط الصيغة فنجد أن حجم العينة حسب التوزيع الأمثل يساوى :

$$n = \frac{\left[\sum_{h=1}^{L} N_h S_h \sqrt{C_h}\right]\left[\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}\right]}{N^2 D + \sum_{h=1}^{L} N_h S_h^2} \quad .... (5 - 51)$$

تطبيق (٥ - ١٠) :

يتكون مجتمع من الموظفين من (٧) موظفين إنفاقهم الشهرى (بآلاف الريالات) كما يلى :

المنطقة (أ) : ٢ ، ٤ ، ٤ ، ١٠

المنطقة (ب) : ١ ، ٢ ، ٦

نريد اختيار عينة حجمها (٥) موظفين وتخصيصها باستخدام طريقة التخصيص الأمثل وذلك إذا كانت تكلفة الوحدة فى الطبقة الأولى $C_1 = 4$ وتكلفة الوحدة فى الطبقة الثانية $C_2 = 9$. المطلوب :

- تحديد حجم العينة فى كل طبقة .
- تقدير متوسط المجتمع .
- إيجاد تباين التخصيص الأمثل للعينة الأولى من العينات الممكنة .

الحـــل :

من بيانات المثال نجد أن :

$N = 7 , N_1 = 4 , N_2 = 3 , n = 5 , C_1 = 4 , C_2 = 9$

كذلك نجد باستخدام العلاقتين التاليتين :

$$\overline{X}_h = \sum_{i=1}^{N_h} X_{hi} / N_h$$

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (X_{hi} - \overline{X}_h)^2$$

أن :

$$\overline{X}_1 = 5 \ , \ \overline{X}_2 = 3 \ , \ S_1^2 = 12 \ , \ S_2^2 = 7$$

ويكون حجم العينة للطبقة الأولى :

$$n_1 = n \frac{N_1 \quad S_1 / \sqrt{C_h}}{\sum_{h=1}^{L} (N_h \ S_h \ / C_h)}$$

$$= 6 \ x \ \frac{4 \ x \ 3.4 \ / \ 2}{\frac{4 \ x \ 3.4}{2} + \frac{3 \ x 2.7}{3}} = \frac{34}{9.5} = 3.5 \approx 4$$

$$n_2 = 5 \ x \ \frac{3 \ x \ 2.7 \ / \ 3}{9.5} = \frac{13.5}{9.5} = 1.4 = 1$$

أى أن توزيع العينة على الطبقات يكون على الشكل التالى (٤ وحدات للطبقة الأولى ووحدة واحدة للطبقة الثانية) :

رقم العينة الممكنة	مفردات الطبقة الأولى	مفردات الطبقة الثانية
١	٢ ، ٤ ، ٤ ، ١٠	١
٢	٢ ، ٤ ، ٤ ، ١٠	٢
٣	٢ ، ٤ ، ٤ ، ١٠	٦

– لتقدير متوسط المجتمع بافتراض أن العينة المختارة هى العينة الممكنة الأولى :

$$(\overline{x}_{opt}) = \frac{\sum_{h=1}^{L} N_h \overline{x}_h}{N}$$

$$= \frac{(4 \ x \ 5) + (3 \ x \ 1)}{7} = \frac{23}{7} = 3.28$$

ويكون تباين تقدير متوسط المجتمع الأمثل باستخدام الصيغة (49 - 5) .

$$V(\overline{x}_{opt}) = \frac{1}{49}\,\frac{1}{5}\left[(4 \times 3.4 \times 2) + (3 \times 2.7 \times 3)\right] \times \left[(4 \times 3.4 / 2) + (3 \times 2.7 / 3)\right] - \frac{1}{49}\left[(4 \times 12) + (3 \times 7)\right]$$

$$= \frac{1}{245}\left[27.2 + 24.3\right] \times \left[6.8 + 2.7\right] - \frac{1}{49}\left[69\right]$$

$$= (\frac{1}{245} \times 51.5 \times 9.5) - 1.4 = 0.60$$

٥ - ٦ - ٤ طريقة نيمان للتخصيص : (Neyman Allocation)

نجد أحيانًا أن تكاليف المعاينة لا تختلف من صيغة لأخرى حيث نجد أن (C_h) متشابهة فى جميع الطبقات . إذا رمزنا للتكلفة (C_h) فى هذه الحالة بالرمز (C_f) ، تصبح دالة التكاليف :

$$C = C_o + C_f \sum_{h=1}^{L} n_h$$

ومنه

$$\boxed{C = C_o + C_f\, n} \quad \text{.... (5 - 52)}$$

من هذه العلاقة نجد حجم العينة (n) باستخدام طريقة نيمان للتخصيص :

$$n = \frac{C - C_o}{C_f}$$

وسنقوم بتقدير حجم العينة حسب طريقة نيمان فيما بعد .

ولتخصيص العينة على الطبقات ، نريد إيجاد (n_h) بحيث يكون $V(\overline{x}_{st})$ أقل ما يمكن باستخدام حجم ثابت للعينة (n) .

لدينا :

$$V(\overline{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h}\,\frac{N_h^2 S_h^2}{n_h}$$

وباستخدام دالة لاغرانج نجد أن :

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \quad .... (5 - 53)$$

بافتراض أن : ثابت = $C_h = C_r$.

تعنى العلاقة السابقة أن حجم الطبقة فى العينة (n_h) يتناسب مع $(N_h S_h)$ أى أن تخصيص العينة على الطبقات يتوقف على حجم الطبقة فى المجتمع ودرجة تجانسها . فإذا كانت الطبقة فى المجتمع كبيرة ، فإننا نسحب منها عينة جزئية كبيرة . كذلك نسحب عينة كبيرة إذا كانت الطبقة فى المجتمع غير متجانسة والعكس بالعكس .

لقد اقترحت هذه الطريقة من قبل (J. Neyman) فى عام ١٩٣٤م وسميت باسمه .

تقديرات طريقة نيمان للتخصيص :

أ - تقدير متوسط المجتمع :

إن مقدر وسطى المجتمع حسب طريقة نيمان :

$$\bar{x}_{Ney} = \bar{x}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{x}_h}{N} \quad .... (5 - 54)$$

وهو مقدر غير متحيز لمتوسط المجتمع .

ب - تباين تقدير متوسط المجتمع :

أما تباين تقدير متوسط المجتمع حسب طريقة نيمان ولنرمز له بالرمز $V(\bar{x}_{Ney})$ فنستطيع حسابه بتبديل قيمة (n_h) فى $V(\bar{x}_{st})$ فيكون لدينا :

$$V(\bar{x}_{Ney}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2 S_h^2}{(nN_h S_h) / \sum_{h=1}^{L} N_h S_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h S_h^2$$

$$= \frac{1}{N^2} \frac{\left[\sum_{h=1}^{L} (N_h S_h)\right]\left[\sum_{h=1}^{L} N_h S_h\right]}{n} - \frac{\sum_{h=1}^{L} N_h S_h^2}{N^2}$$

ومنه نجد أن :

$$V(\overline{x}_{Ney}) = \frac{1}{N^2}\frac{(\sum_{h=1}^{L} N_h S_h)^2}{n} \quad \frac{\sum_{h=1}^{L} N_h S_h^2}{N^2} \qquad \text{.... (5 - 55)}$$

وعند عدم معرفة S_h^2 نستخدم تقديرها من عينة (s_h^2) ونحصل على تقدير تباين $(\overline{x}_{st})$ ولنرمز له بالرمز $\hat{V}(\overline{x}_{Ney})$.

جـ - تحديد حجم العينة الطبقية حسب طريقة نيمان للتخصيص :

لدينا :

$$B^2 = Z^2 \, V(\overline{x}_{st})$$

ومنه

$$V(\overline{x}_{st}) = \frac{B^2}{Z^2} = D$$

$$V(\overline{x}_{Ney}) = \frac{1}{N^2}\frac{(\sum_{h=1}^{L} N_h S_h)^2}{n} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2$$

ومنه نجد أن :

$$N^2 D + \sum_{h=1}^{L} N_h S_h^2 = (\sum_{h=1}^{L} N_h S_h)^2 / n$$

$$n = \frac{(\sum_{h=1}^{L} N_h S_h)^2}{N^2 D + \sum_{h=1}^{L} N_h S_h^2} \qquad \text{.... (5 - 56)}$$

حيث $D = \frac{B^2}{Z^2}$, B حد الخطأ المقبول وعند عدم معرفة (S_h^2) نستخدم تقديرها من عينة (s_h^2) .

تطبيق (٥ – ١١) :

باستخدام بيانات المثال (٥ – ١٠) ، أوجد حجم العينة لكل طبقة ، ثم احسب تقدير متوسط المجتمع وتباينه باستخدام طريقة نيمان للتخصيص .

الحل :

لدينا

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} \cdot n$$

$$n_1 = \frac{4 \times 3.4}{(4 \times 3.4) + (3 \times 2.7)} \times 5 = \frac{13.6}{13.6 + 8.1} \times 5 = \frac{68.0}{21.7} = 3.1 \approx 3$$

$$n_2 = \frac{3 \times 2.7}{(4 \times 3.4) + (3 \times 2.7)} \times 5 = \frac{40.5}{21.7} = 1.9 \approx 2$$

إذا افترضنا أن مفردات العينة المختارة كانت (٢ ، ٤ ، ٤) من الطبقة الأولى و (١ ، ٢) من الطبقة الثانية ، يكون :

$$\bar{x}_1 = 3.3 \qquad \bar{x}_2 = 1.5$$

وبالتالى يكون تقدير متوسط المجتمع :

$$\bar{x}_{st} = \bar{x}_{Ney} = \frac{(4 \times 3.4) + (3 \times 1.5)}{7}$$

$$= \frac{18.1}{7} = 2.58$$

$$V(\bar{x}_{Ney}) = \frac{1}{49} \times \frac{(21.7)^2}{5} - \frac{1}{49}\left[(4 \times 12) + (3 \times 7)\right]$$

$$= \frac{470.9}{245} - \frac{69}{49} = 1.92 - 1.41 = 0.51$$

تطبيق (٥ – ١٢) :

مجتمع من الأشخاص دخولهم الشهرية موزعة على طبقتين ، وكانت لدينا البيانات التالية :

$$N = 10 \; , \; N_1 = 6 \; , \; N_2 = 4 \; , \; S_1^2 = 16 \; , \; S_2^2 = 9 \; , \; C_1 - 9 \; , \; C_2 = 4 \; , \; L = 2$$

سحبنا عينة طبقية من هذا المجتمع وكان الخطأ المسموح به (٢) واحتمال الحصول على الدقة (٩٥٪) المطلوب تحديد حجم العينة حسب طرق التخصيص التالية :

- طريقة التخصيص المتساوى .
- طريقة التخصيص المتناسب .
- طريقة التخصيص الأمثل .
- طريقة نيمان للتخصيص .

الحـل :

نضع البيانات التالية التى تساعدنا فى تحديد حجم العينة :

N_1	N_2	S_1	S_2	S_1^2	S_2^2	N_1S_1	N_2S_2	$N_1S_1^2$	$N_2S_2^2$	C_1	C_2
6	4	4	3	16	9	24	12	96	36	9	4

كذلك نجد أن :

$$\sigma_1^2 = \frac{N_1 - 1}{N_1} S_1^2 = \frac{6-1}{6} \times 16 = 13.33 \; , \; \sigma_1 = 3.65$$

$$\sigma_2^2 = \frac{N_2 - 1}{N_2} S_2^2 = \frac{4-1}{4} \times 9 = 6.75 \; , \; \sigma_2 = 2.60$$

$$\sum_{h=1}^{L} N_h S_h = N_1 S_1 + N_2 S_2 = 24 + 12 = 36$$

$$\sum_{h=1}^{L} N_h S_h^2 = N_1 S_1^2 + N_2 S_2^2 = 96 + 36 = 132$$

$$\sqrt{C_1} = 3 \; , \; \sqrt{C_2} = 2 \; , \; N^2 = 100$$

$$\sum_{h=1}^{L} N_h^2 S_h^2 = N_1^2 S_1^2 + N_2^2 S_2^2 = (36 \times 16) + (16 \times 9) = 720$$

$$D = \frac{B^2}{Z^2} = \frac{(2)^2}{(1.96)^2} = \frac{4}{3.84} = 1.04$$

$$W_1 = \frac{N_1}{N} = 0.6 \text{ , } W_2 = \frac{N_2}{N} = 0.4$$

١ - تحديد حجم العينة باستخدام طريقة التخصيص المتساوى :

بالتعويض فى الصيغة (45 - 5) نجد أن :

$$n = \frac{2 \times 720}{100 \times 1.04 + 132} = \frac{1440}{104 + 132} = \frac{1440}{236}$$

$$= 6.1 \quad \approx 6$$

أى أن حجم العينة هو ستة أشخاص وحجم كل طبقة (٣) أشخاص .

$$n_1' = n_2 = \frac{n}{L} = \frac{6}{2} = 3$$

٢ - تحديد حجم العينة باستخدام طريقة التخصيص المتناسب :

نعوض فى الصيغة (43 - 5) فنجد أن :

$$n = \frac{N \sum_{h=1}^{L} N_h S_h^2}{N^2 D + \sum_{h=1}^{L} N_h S_h^2}$$

$$= \frac{10 \times 132}{104 + 132} = \frac{1320}{236} = 5.59$$

$$= 6$$

أى أن حجم العينة حسب طريقة التخصيص المتناسب هو ستة أشخاص ، ويكون حجم الطبقة الأولى وحجم الطبقة الثانية على التوالى :

$$n_1 = n \frac{N_1}{N} = 6 \times \frac{6}{10} = 3.6 \approx 4$$

$$n_2 = n \frac{N_2}{N} = 6 \times \frac{4}{10} = 2.4 \approx 2$$

٣ – تحديد حجم العينة باستخدام طريقة التخصيص الأمثل :

نطبق الصيغة (51 - 5) فنجد أن :

$$n = \frac{\left[(6 \times 4 \times 3) + (4 \times 3 \times 2)\right]\left[(24/3) + (12/2)\right]}{(100 \times 1.04) + 132} = \frac{96 \times 14}{236}$$

$$= \frac{1344}{236} = 5.69 \approx 6$$

أى أن حجم العينة حسب طريقة التخصيص الأمثل هو ستة أشخاص ويتم التوزيع على الطبقتين باستخدام الصيغة (47 - 5) كما يلى :

$$n_1 = 6 \frac{24/3}{(24/3) + (12/2)}$$

$$= 6 \times \frac{8}{14} = \frac{48}{14} = 3.43 \approx 3$$

$$n_2 \doteq 6 \frac{12/2}{14} = \frac{36}{14} = 2.57 \approx 3$$

٤ – تحديد حجم العينة باستخدام طريقة نيمان للتخصيص :

نستخدم الصيغة (56 - 5) فنجد أن :

$$n = \frac{(36)^2}{(100 \times 1.04) + 132} = \frac{1296}{236}$$

$$= 5.49 \approx 5$$

ويكون حجم الطبقة الأولى باستخدام الصيغة (53 - 5) :

$$n_1 = 5 \times \frac{24}{24 + 12} = 3.33 \approx 3$$

وحجم الطبقة الثانية :

$$n_2 = 5 \times \frac{12}{36} = 1.67 \approx 2$$

٥ - ٧ المقارنة بين المعاينة العشوائية البسيطة والمعاينة الطبقية العشوائية :

٥ - ٧ - ١ المقارنة بين المعاينة العشوائية البسيطة والمعاينة الطبقية العشوائية بطريقة التخصيص المتناسب والأمثل :

بعد أن درسنا المعاينة العشوائية البسيطة والمعاينة الطبقية العشوائية ، قد نتساءل : هل المعاينة الطبقية العشوائية أدقّ وأكفأ من المعاينة العشوائية البسيطة :

إذا رمزنا إلى تباين المعاينة العشوائية البسيطة بالرمز $V(\overline{x}_{ran})$ وتجاهلنا $\frac{1}{N_h}$ في المجتمعات الكبيرة ، يمكننا القول إن :

$$V(\overline{x}_{opt}) \leq V(\overline{x}_{prop}) \leq V(\overline{x}_{ran})$$

أى أن مقدرات المعاينة الطبقية العشوائية هي أكفأ عادة من مقدرات المعاينة العشوائية البسيطة . للبرهان على ذلك ، نعلم من تحليل تباين المجتمع الطبقي أن :

$$S^2 = \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} (x_{hi} - \overline{X})^2$$

وبإضافة وطرح $(\overline{X}_h)$ نجد أن :

$$(N-1) S^2 = \sum_{h=1}^{L} \sum_{i=1}^{N_h} (x_{hi} - \overline{X}_h)^2 + \sum_{h=1}^{L} N_h (\overline{X}_h - \overline{X})^2$$

$$= \sum_{h=1}^{L} (N_h - 1) S_h^2 + \sum_{h=1}^{L} N_h (\overline{X}_h - \overline{X})^2$$

وبإهمال $\frac{1}{N_h}$ - و $\frac{1}{N}$ نجد أن :

$$S^2 = \sum W_h S_h^2 + \sum W_h (\overline{X}_h - \overline{X})^2$$

أى أن :

$$V(\overline{x}_{ran}) = (1-f) \frac{S^2}{n}$$

$$= \frac{1-f}{n} \sum_{h=1}^{L} W_h S_h^2 + \frac{1-f}{n} \sum_{h=1}^{L} W_h (\overline{X}_h - \overline{X})^2$$

$$= V(\overline{x}_{prop}) + \frac{1-f}{n} \sum_{h=1}^{L} W_h (\overline{X}_h - \overline{X})^2$$

وهكذا نجد أن تباين العينة العشوائية البسيطة أكبر من تباين العينة الطبقية بطريقة التخصيص المتناسب بالمقدار* :

$$V(\bar{x}_{ran}) - V(\bar{x}_{prop}) = \frac{1-f}{n} \sum_{h=1}^{L} W_h (\bar{X}_h - \bar{X})^2$$

ونعلم أن تباين العينة الطبقية حسب طريقة التخصيص المتناسب أكبر من هذا التباين بطريقة التخصيص الأمثل بالمقدار :

$$V(\bar{x}_{prop}) - V(\bar{x}_{opt}) = \frac{1}{n} \left[\sum W_h S_h^2 - \left(\sum W_h S_h \right)^2 \right]$$

وهكذا نجد بإهمال $\frac{1}{N_h}$ أن :

$$V(\bar{x}_{ran}) = V(\bar{x}_{opt}) + \frac{1}{n} \left[\sum_{h=1}^{L} W_h S_h^2 - \left(\sum_{h=1}^{L} W_h S_h \right)^2 \right] + \frac{(1-f)}{n} \sum_{h=1}^{L} W_h (\bar{X}_h - \bar{X})^2 \quad .... (5-57)$$

ويلاحظ أن هناك مقدارين يمثلان الفرق بين العينة العشوائية البسيطة والعينة الطبقية حسب التوزيع الأمثل :

- المقدار الأخير يمثل الزيادة التى حصلت نتيجة حذف الفروق بين متوسطات الطبقات .
- المقدار الثانى (الأوسط) يمثل الفروق التى نتجت بسبب الفروق بين الانحرافات المعيارية للطبقات . ويمثل هذا المقدار الفرق بين تباين العينة الطبقية حسب طريقة التخصيص الأمثل وحسب طريقة التخصيص المتناسب .

وباستبدال (S^2) بقيمتها وإدخال $\frac{1}{N_h}$ نجد أن :

$$V(\bar{x}_{ran}) = V(\bar{x}_{prop}) + \frac{1-f}{n\,(N-1)} \left[\sum_{h=1}^{L} N_h (\bar{X}_h - \bar{X})^2 - \frac{1}{N} \sum_{h=1}^{L} (N - N_h) S_h^2 \right] .... (5-58)$$

---

* Cochran W. : Sampling Techniques, John Wiley & Sons, New York, 1977 (p.p. 99 - 101) .

ويمكننا القول إن العينة الطبقية بطريقة التخصيص المتناسب تعطى عادة تباينًا أقل أى أن :

$$V(\overline{x}_{ran}) \geqslant V(\overline{x}_{prop})$$

أى أن التقديرات من بيانات عينة طبقية حسب طريقة التخصيص المتناسب أكفأ عادة من التقديرات من بيانات عينة عشوائية بسيطة .

٥ - ٧ - ٢ مقارنة دقة المعاينة العشوائية البسيطة والمعاينة الطبقية بطريقة التخصيص المتناسب وتخصيص نيمان :

نعلم أن تباين العينة الطبقية حسب طريقة التخصيص المتناسب وطريقة نيمان للتخصيص يساوى :

$$V(\overline{x}_{prop}) = \frac{1}{nN}\sum_{h=1}^{L} N_h S_h^2 - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2$$

$$V(\overline{x}_{Ney}) = \frac{1}{N^2}\,\frac{(\sum_{h=1}^{L} N_h S_h)^2}{n} - \frac{1}{N^2}\sum_{h=1}^{L} N_h S_h^2$$

لنوضح الفرق بين المقدّرين وأيّهما أكفأ . نقول إن الفرق بينهما يساوى :

$$V(\overline{x}_{prop}) - V(\overline{x}_{Ney}) = \frac{\sum_{h=1}^{L} N_h S_h^2}{N\,n} - \frac{(\sum_{h=1}^{L} N_h S_h)^2}{n\,N^2}$$

$$= \frac{1}{n\,N}\left[\sum_{h=1}^{L} N_h S_h^2 - \frac{(\sum_{h=1}^{L} N_h S_h)^2}{N}\right]$$

$$= \frac{1}{n\,N}\left[\sum_{h=1}^{L} N_h S_h^2 - \frac{2}{N}(\sum_{h=1}^{L} N_h S_h)^2 + \frac{\sum_{h=1}^{L} N_h}{N^2}(\sum_{h=1}^{L} N_h S_h)^2\right]$$

حيث يساوى الحدان الأخيران من الصيغة السابقة المقدار :

$$- \frac{(\sum_{h=1}^{L} N_h S_h^2)}{N}$$

أى أن الفرق بين تباين العينة حسب طريقة التخصيص المتناسب وحسب طريقة نيمان يساوى :

$$V(\overline{x}_{prop}) - V(\overline{x}_{Ney}) = \frac{1}{nN} \sum_{h=1}^{L} N_L [S_h^2 - \frac{2}{N} S_h (\sum_{h=1}^{L} N_h S_h) + \frac{1}{N^2} (\sum_{h=1}^{L} N_h S_h)^2]$$

$$= \frac{1}{nN} \sum_{h=1}^{L} N_h (S_h - \frac{1}{N} \sum_{h=1}^{L} N_h S_h)^2$$

فإذا رمزنا إلى المقدار $\frac{1}{N} \sum N_h S_h$ بالرمز $(\overline{S})$ يكون :

$$V(\overline{x}_{prop}) - V(\overline{x}_{Ney}) = \frac{1}{nN} \sum_{h=1}^{L} N_h (S_h - \overline{S})^2$$

أى أن :

$$V(\overline{x}_{prop}) = V(\overline{x}_{Ney}) + \frac{1}{nN} \sum_{h=1}^{L} N_h (S_h - \overline{S})^2 \qquad .... (5 - 59)$$

حيث

$$\overline{S} = \frac{1}{N} \sum_{h=1}^{L} N_h S_h$$

يتضح من الصيغة (5 - 59) أن تباين تقدير متوسط عينة طبقية بطريقة نيمان للتخصيص (أو حتى حسب طريقة التخصيص الأمثل لأن طريقة نيمان هى حالة خاصة من طريقة التخصيص الأمثل حيث تكون التكلفة متشابهة بين جميع الطبقات) يكون دائماً أقل أو يساوى تباين العينة الطبقية بطريقة التخصيص المتناسب ، أى أن :

$$V(\overline{x}_{opt}) \leq V(\overline{x}_{prop})$$

$$V(\overline{x}_{Ney}) \leq V(\overline{x}_{prop})$$

وتتوقف قيمة الفرق بينهما على مقدار الانحرافات المعيارية للطبقات أى تتوقف على (S_h) .

ويمكننا القول إنه عمليًا ، إذا كان هناك اختلافات كبيرة فى الانحرافات المعيارية بين الطبقات ، فإن استخدام توزيع نيمان يكون الأفضل . وبالعكس إذا كانت هذه الاختلافات غير كبيرة فإن استخدام التوزيع المتناسب يكون الأفضل .

ويتضح من الصيغة (5 - 59) أن مقدرات طريقة تخصيص نيمان أكفأ من مقدرات طريقة التخصيص المتناسب لأن تباين طريقة نيمان للتخصيص أصغر من تباين طريقة التخصيص المتناسب .

ويمكننا القول إن :

$$V(\overline{x}_{opt}) \leq V(\overline{x}_{Ney}) \leq V(\overline{x}_{prop}) \leq V(\overline{x}_{ran}) \qquad .... (5 - 60)$$

وذلك لأنه كلما اختلفت متوسطات الطبقات ، نحصل عادة على دقة أكبر باستخدام العينة الطبقية المتناسبة على العينة العشوائية البسيطة .

كذلك كلما ازدادت الفروق بين الانحرافات المعيارية للطبقات (S_h) كلما ازدادت الدقة باستخدام طريقة التخصيص الأمثل عما لو استخدمنا طريقة التخصيص المتناسب ، ونستطيع كتابة الصيغة التالية (باستخدام الصيغتين الأخيرتين) :

$$V(\overline{x}_{ran}) = V(\overline{x}_{Ney}) + \frac{\sum N_h (S_h - \overline{S})^2}{n\,N} + \frac{\sum N_h (\overline{x}_h - \overline{\overline{X}})^2}{N\,n}$$

أى أن مقدار ما نحصل عليه من كسب فى الدقة الذى ينتج باستخدام طريقة نيمان للتخصيص عوضًا عن استخدام المعاينة العشوائية البسيطة ينتج من عاملين :

١ – الفروق بين متوسطات الطبقات .

٢ – الفروق بين الانحرافات المعيارية للطبقات .

لذا فإن تقسيم المجتمع إلى طبقات واختيار طريقة التخصيص المناسبة ، يجب أن يتم بدقة ، وذلك للحصول على النتائج المطلوبة والدقيقة بأسهل الطرق وأفضلها .

تطبيق (٥ – ١٣) :

يتكون مجتمع من الموظفين من طبقتين سنوات خبراتهم كما يلى :

الطبقة الأولى : ٢ ، ٤ ، ٦

الطبقة الثانية : ٩ ، ١٢ ، ١٨ ، ٢١

المطلوب :

أ – حساب كل من إجمالى سنوات الخبرة ومتوسط سنوات الخبرة لكل طبقة وإجمالى سنوات الخبرة للموظفين .

ب – اسحب عينة من حجم $n_1 = 2 , n_2 = 2$ وقدر المعالم المطلوبة فى (أ) باستخدام العينة الأولى واذكر جميع العينات الممكن سحبها .

ج – تقدير تباين تقدير متوسط العينة باستخدام بيانات العينة الأولى .

د – قدر متوسط سنوات الخبرة بمستوى ثقة (٩٥٪) .

هـ – ما هو حجم العينة المناسب لاختيار عينة من الموظفين من مجتمع حجمه (١٠٠) موظف موزّعين على طبقتين بالتساوى إذا كان خطأ التقدير المطلوب (١,٥) سنة .

الحـل :

أ – نستخدم الصيغة التالية لاستخراج سنوات الخبرة للطبقة (h) :

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

ويكون

$$X_1 = 2 + 4 + 6 = 12$$
$$X_2 = 9 + 12 + 18 + 21 = 60$$

ومتوسط سنوات الخبرة فى الطبقة (h) :

$$\overline{X}_h = \sum_{i=1}^{N_h} X_{hi} / N_h$$

ويكون متوسط سنوات الخبرة للطبقة الأولى :

$$\overline{X}_1 = 12 / 3 = 4$$

ومتوسط سنوات الخبرة للطبقة الثانية :

$$\overline{X}_2 = 60 / 4 = 15$$

إجمالى سنوات الخبرة لجميع الموظفين :

$$X = \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}$$

$$= 12 + 60 = 72$$

ويكون متوسط سنوات الخبرة لجميع الموظفين :

$$\overline{X} = \sum_{h=1}^{L} N_h \ \overline{X}_h \ / N_h$$

$$= \left[(3 \times 4) + (4 \times 15) \right] / 7$$

$$= 10.285$$

أى (10.29) سنة .

ب – إن عدد العينات الممكن سحبها هو :

$$\binom{N_1}{n_1} \binom{N_2}{n_2} = \binom{3}{2} \binom{4}{2} = 18$$

وهذه العينات هى :

المفردات	رقم العينة	المفردات	رقم العينة
2,6,12,18	10	2,4,9,12	1
2,6,12,21	11	2,4,9,18	2
2,6,18,21	12	2,4,9,21	3
4,6,9,12	13	2,4,12,18	4
4,6,9,18	14	2,4,12,21	5
4,6,9,21	15	2,4,18,21	6
4,6,12,18	16	2,6,9,12	7
4,6,12,21	17	2,6,9,18	8
4,6,18,21	18	2,6,9,21	9

لتقدير متوسط سنوات الخبرة نفترض أن العينة التى تم سحبها هى 2,4,9,12 أى العينة الأولى ويكون لدينا باستخدام الصيغة التالية :

$$\overline{x}_h = \sum_{i=1}^{n_h} x_{hi} / n_h$$

$$\widehat{X}_1 = \bar{x}_1 = (2 + 4) / 2 = 3$$

$$\widehat{X}_2 = \bar{x}_2 = (9 + 12) / 2 = 10.5$$

وبالتالى يصبح تقدير متوسط سنوات الخبرة لدى الموظفين :

$$\widehat{X} = \bar{x}_{st} = \sum_{h=1}^{L} N_h \bar{x}_h / N$$

$$\bar{x}_{st} = \left[(3x3) + (4 \text{ x } 10.5)\right] / 7$$

$$= \frac{51}{7} = 7.28$$

أى أن تقدير متوسط سنوات الخبرة للموظفين يساوى (7.28) سنة .

– تقدير تباين تقدير متوسط المجتمع :

نستخدم الصيغة التالية :

$$\hat{V}(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h} - \frac{1}{N^2} \sum_{h=1}^{L} N_h s_h^2$$

نعلم أن :

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$$

ويكون هذا التباين للعينة الطبقية الأولى الممكن سحبها : .

$$s_1^2 = \frac{1}{2-1} [(2-3)^2 + (4-3)^2] = 2$$

$$s_2^2 = \frac{1}{2-1} [(9-10.5)^2 + (12-10.5)^2] = 4.5$$

$$\hat{V}(\bar{x}_{st}) = \frac{1}{7^2} \left[\frac{3^2 \text{ x } 2}{2} + \frac{4^2 \text{ x } 4.5}{2}\right] - \frac{1}{7^2} \left[(3 \text{ x } 2) + (4 \text{ x } 4.5)\right]$$

$$= 0.917 - 0.489 = 0.428$$

د – حدا الثقة هما :

$$\bar{x}_{st} + t_{(1-\alpha/2,\, n-1)}\, \sigma_{\bar{x}_{st}}$$

أى :

$$7.28 \mp 2.3534 \sqrt{0.428}$$

أى يساوى 5.741 و 8.819 .

أى أن متوسط سنوات الخبرة يتراوح بين 5.741 سنة و 8.819 بمستوى ثقة (95%) .

هـ – تحديد حجم العينة المناسب حسب طريقة التخصيص المتساوى :

$$n = \frac{L \sum_{h=1}^{L} N_h^2\ s_h^2}{N^2 D + \sum_{h=1}^{L} N_h\ s_h^2}$$

$$D = \frac{\beta^2}{4} = \frac{(1.5)^2}{4} = 0.56$$

$$n = \frac{2\,[50^2 \times 2 + 50^2 \times 4.5]}{(100^2 \times 0.56) + [50 \times 2) + (50 \times 4.5)]}$$

$$= \frac{32500}{5925} = 5.48 \approx 6$$

ويكون

$$n_1 = n_2 = 3$$

الفصل
السادس

المعاينة الطبقية للنسب

(Stratified Sampling of Proportions)

٦ - ١ رموز وتعاريف :

تتطلب كثير من التطبيقات العملية ، تقدير نسبة المفردات التى تتصف بخاصية معينة فى مجتمع مقسم إلى طبقات مثلاً . قد نرغب فى تقدير نسبة الموظفين الذين التحقوا بدورات تدريبية فى الجهات الحكومية حسب التخصص أو حسب نوع الإدارة (العليا ، المتوسطة ، التنفيذية) . تسمى المعاينة التى نستخدمها فى هذه الحالات المعاينة الطبقية للنسب ، ويستخدم هذا النوع من المعاينات فى كثير من البحوث الإدارية والاقتصادية التى تهدف إلى تقدير نسب الوحدات التى تتصف بخاصية معينة كبحوث الرضا الوظيفى والدخل القومى والعمالة ومراقبة جودة الإنتاج وغيرها .

نعلم مما سبق أن (X_{hi}) تمثل المفردة (i) فى الطبقة (h) حيث لدينا (L) طبقة (أى $h = 1,2,\ldots, L$) . وأن متوسط الطبقة (h) فى المجتمع .

$$\overline{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

حيث (N_h) يمثل حجم الطبقة (h) فى المجتمع .

كذلك وجدنا أن متوسط المجتمع $(\overline{X})$ المقسم إلى طبقات يساوى :

$$\overline{X} = \frac{1}{N} \sum_{h=1}^{L} N_h \overline{X}_h$$

وعند استخدام معاينة النسب ، نجد أن $(X_{hi}=1)$.

عندما تتصف الوحدة بالخاصية المطلوبة و $(X_{hi}=0)$ عندما لا تتصف الوحدة بالخاصية المطلوبة .

إذا رمزنا للنسبة (P_h) للتعبير عن نسبة المجتمع فى الطبقة (h) لخاصية معينة ، نجد أن

$$P_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi} \qquad \ldots (6 - 1)$$

حيث $(i=1,2,\ldots,N_h)$ ، أى أن : $P_h = \overline{X}_h$

وتكون نسبة المجتمع ولنرمز لها بالرمز (P) :

$$P = \frac{1}{N} \sum_{h=1}^{L} N_h P_h \quad .... (6 - 2)$$

وتساوى هذه النسبة :

$$P = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi} \quad .... (6 - 3)$$

حيث (X_{hi}) تساوى (١) أو (0) .

٦ - ٢ تقدير نسبة المجتمع :

تكون قيم مفردات المجتمع فى كثير من الأحيان مجهولة ، لذا يتم استخدام عينة طبقية لتقدير نسبة المجتمع . نعلم أن متوسط العينة الطبقية ($\overline{x}_{st}$) يساوى

$$\overline{x}_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h \overline{x}_h \quad .... (6 - 4)$$

حيث

$$\overline{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$$

وباستخدام النسب نجد أن المقدر المستخدم لتقدير نسبة المجتمع لعينة طبقية يساوى :

$$\hat{P} = p_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h p_h \quad .... (6 - 5)$$

حيث

$$p_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \qquad \text{.... (6 - 6)}$$

و (x_{hi}) تساوى الواحد عندما تتصف الوحدة بالخاصية وتساوى الصفر عندما لا تتصف بذلك :

إن (p_{st}) هو مقدر غير متحيز لـ (P) وذلك لأن :

$$E(p_{st}) = \frac{1}{N} E\left(\sum_{h=1}^{L} N_h p_h \right)$$

$$= \frac{1}{N} \sum_{h=1}^{L} N_h E(p_h)$$

$$= \frac{1}{N} \sum_{h=1}^{L} N_h \frac{1}{n_h} \sum_{i=1}^{n_h} E(x_{hi})$$

$$= \frac{1}{N} \sum_{h=1}^{L} N_h \frac{1}{n_h} n_h \sum_{i=1}^{N_h} x_{hi} \frac{1}{N_h}$$

حيث $\frac{1}{N_h}$ هو احتمال الحصول على المفردة x_{hi} من الطبقة (h)

$$E(p_{st}) = \frac{1}{N} \sum_{h=1}^{L} N_h \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$$

$$= \frac{1}{N} \sum_{h=1}^{L} N_h P_h = P$$

حيث

$$P_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$$

أى أن : $E(p_{st}) = P$

أى أن (p_{st}) هو مقدر غير متحيز لنسبة المجتمع (P) .

ويمكننا القول إن نسبة الذين لا يتصفون بالخاصية فى الطبقة (h) لمجتمع مقسم إلى طبقات يساوى :

$$Q_h = 1 - P_h \quad .... (6 - 7)$$

وتساوى هذه النسبة فى المجتمع :

$$Q = 1 - P \quad .... (6 - 8)$$

ومقدر نسبة المجتمع من بيانات عينة طبقية للذين لا يتصفون بالخاصية هو :

$$\widehat{Q} = 1 - \widehat{P} \quad .... (6 - 9)$$

$$\widehat{Q}_{st} = 1 - \widehat{P}_{st} \quad .... (6 - 10)$$

ومقدر نسبة المجتمع من بيانات عينة طبقية للذين لا يتصفون بالخاصية فى الطبقة (h) هو :

$$q_h = 1 - p_h \quad .... (6 - 11)$$

٦ - ٢ تباين التقديرات للمعاينة الطبقية للنسب وتقديراتها :

٦ - ٢ - ١ تباين تقدير نسبة المجتمع :

نعلم أن تباين نسبة المجتمع عندما تأخذ (x_i) القيم الصفر أو الواحد يساوى :

$$S^2 = \frac{N}{N - 1} P Q \quad .... (6 - 12)$$

كذلك نعلم أن تباين النسبة للطبقة (h) يساوى :

$$S_h^2 = \frac{N_h}{N_h - 1} P_h Q_h \quad .... (6 - 13)$$

وأن تباين ($\bar{x}_{st}$) هو تباين (p_{st}) أى أن :

$$V(\bar{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 S_h^2}{n_h} \quad .... (6 - 14)$$

وعندما تأخذ (x_i) القيم الصفر أو الواحد نجد أن :

$$V(P_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \times \frac{N_h^2}{n_h} \times \frac{N_h}{N_h - 1} P_h Q_h \quad .... (6 - 15)$$

وبالاختصار نجد أن :

$$V(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{P_h Q_h}{n_h} \quad .... (6 - 16)$$

وعندما يكون المجتمع كبيراً ومعامل تصحيح المجتمع المحدود يساوى الواحد تصبح الصيغة (6 - 15) :

$$V(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{P_h Q_h}{n_h} \quad .... (6 - 17)$$

٦ - ٣ - ٢ تقدير تباين تقدير نسبة المجتمع :

نجد فى التطبيقات العملية أن نسبة المجتمع للطبقة (h) غير معلومة ، لذا نقدرها من بيانات عينة ويصبح مقدر تباين النسبة للعينة الطبقية :

$$\hat{V}(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{p_h q_h}{n_h} \quad .... (6 - 18)$$

حيث p_h هو نسبة الطبقة (h) من بيانات العينة . وعندما يكون المجتمع كبيراً ومعامل تصحيح المجتمع المحدود مساوياً للواحد نجد أن :

$$\hat{V}(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{p_h q_h}{n_h} \qquad \text{.... (6 - 19)}$$

إن مقدر تباين تقدير نسبة المجتمع $\hat{V}(p_{st})$ هو مقدر غير متحيز لتباين تقدير نسبة المجتمع $V(p_{st})$ (كما يتضح فى الملحق رقم (٥ - ٢) فى نهاية الكتاب) .

تطبيق (٦ - ١) :

أرادت إحدى المؤسسات دراسة نسبة إنتاجها الردىء لتحسينه . إن $(x_i=1)$ عندما تكون الوحدة معيبة و$(x_i=0)$ عندما تكون الوحدة جيدة ، وكان لدى المؤسسة مجموعتان من الآلات تنتج كل منها خمس وحدات :

المجموعة الأولى :

$$X_{11} = 1 , X_{12} = 0 , X_{13} = 1 , X_{14} = 0 , X_{15} = 0$$

المجموعة الثانية :

$$X_{21} = 1 , X_{22} = 0 , X_{23} = 1 , X_{24} = 1 , X_{25} = 1$$

المطلوب :

١ - استخراج قيمة نسبة الوحدات المعيبة للمجتمع لكل طبقة ثم قيمة هذه النسبة للمجتمع كله .

٢ - استخراج تباين نسبة المجتمع إذا سحبنا عينة حجمها خمس وحدات $(n_1=2 , n_2=3)$.

٣ - تقدير نسبة المجتمع وتباينها إذا كانت قيم مفردات العينة المختارة :

$$x_{11} = 0\ x_{12} = 1 ,\ x_{21} = 0 ,\ x_{22} = 1 ,\ x_{23} = 1$$

الحل :

إن نسبة المجتمع للطبقة الأولى والطبقة الثانية هما :

$$P_1 = \frac{\sum_{i=1}^{N_1} X_{1i}}{N_1} = \frac{1+0+1+0+0}{5}$$

$$= \frac{2}{5} = 0.40$$

$$P_2 = \frac{\sum_{i=1}^{N_2} X_{2i}}{N_2} = \frac{1 + 0 + 1 + 1 + 1}{5}$$

$$= \frac{4}{5} = 0.80$$

وتكون نسبة المجتمع أى نسبة الوحدات المعيبة لإنتاج الآلات :

$$P = \frac{\sum_{h=1}^{L} N_h P_h}{N} = \frac{N_1 P_1 + N_2 P_2}{N}$$

$$= \frac{(5 \times 0.40) + (5 \times 0.80)}{10}$$

$$= \frac{6}{10} = 0.60$$

أى نسبة الوحدات المعيبة فى المؤسسة (٦٠٪) .

٢ – لاستخراج تباين نسبة المجتمع ، نستخدم الصيغة (16 - 6) :

$$V(P_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{P_h Q_h}{n_h}$$

$$= \frac{1}{100} \left[\left(\frac{(5 - 2)}{4} \times 25 \times \frac{(0.4 \times 0.6)}{2} \right) + \left(\frac{(5 - 3)}{4} \times 25 \times \frac{(0.8 \times 0.2)}{3} \right) \right]$$

$$= \frac{1}{100} [2.25 + 0.66]$$

$$= \frac{2.91}{100}$$

$= 0.0291$ أي 2.91%

٣ – لاستخراج تقدير تباين النسبة أى $\hat{V}(p_{st})$، نستخدم إحدى العينات الممكن سحبها (وهى العينة التى تنتج بالاختيار العشوائى من كل طبقة) ولتكن هذه العينة حسب البيانات :

$x_{11} = 0 \; x_{12} = 1 , \; x_{21} = 0 , \; x_{22} = 1 , \; x_{23} = 1$

ولنستخرج النسب التالية :

$$p_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} = \frac{0+1}{2} = 1/2 = 0.50$$

$$p_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2} = \frac{0+1+1}{2} = \frac{2}{3} = 0.666$$

$$p_{st} = \frac{\sum_{h=1}^{L} N_h\, p_h}{N} = \frac{(5 \times 0.5) + (5 \times 0.666)}{10}$$

$$= \frac{5.83}{10} = 0.583 \text{ أي } 58.3\ \%$$

أى أن تقدير نسبة المجتمع تساوى (٥٨,٣٪) . ونستخدم الصيغة التالية لاستخراج تقدير تباين تقدير نسبة المجتمع :

$$\hat{V}(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{p_h\, q_h}{n_h}$$

$$= \frac{1}{100} \left[\left[\frac{(5-2)}{4} \times 25 \times 0.50 \times \frac{0.5}{2} \right] + \left[\frac{5-3}{4} \times 25 \times 0.666 \times \frac{0.334}{3} \right] \right]$$

$$= \frac{1}{100} [2.344 + 0.926] = \frac{3.270}{100}$$

$$= 0.0327$$

٦ - ٤ حدود الثقة لتقديرات نسبة المجتمع والقيمة الكلية للمجتمع :

٦ - ٤ - ١ حدا الثقة لتقدير نسبة المجتمع :

قمنا فيما سبق بتقدير نسبة المجتمع من بيانات عينة طبقية ، حيث توصلنا إلى الصيغة (5 - 6) . ولإيجاد حدى الثقة لنسبة المجتمع ، نستخدم العلاقة التالية :

$$p_{st} - \beta \leq P \leq p_{st} + \beta \qquad \text{.... (6 - 20)}$$

حيث (β) هى حد خطأ التقدير وتساوى :

أ – فى حالة العينات الكبيرة :

$$\beta = Z_{(1 - \alpha/2)} \sqrt{V(p_{st})} \qquad \text{.... (6 - 21)}$$

حيث $V(p_{st})$ هى تباين النسبة وقد نستخدم تقديره عندما يكون مجهولاً أى نستخدم $\hat{V}(p_{st})$ كما سيتضح فيما بعد .

ب – فى حالة العينات الصغيرة :

$$\beta = t_{(1 - \alpha/2,\, n-1)} \sqrt{V(p_{st})} \qquad \text{.... (6 - 22)}$$

حيث (t) هى القيمة الجدولية لتوزيع ستودنت بمستوى ثقة % $\left(1 - \frac{\alpha}{2}\right)$ ودرجات حرية (n-1) .

ويلاحظ أننا نحتاج إلى استخراج قيمة $V(p_{st})$ أو تقديرها كما يتضح فى الصفحات القادمة وذلك لاستخراج حدى الثقة .

٦ - ٤ - ٢ تقدير القيمة الكلية باستخدام تقدير نسبة المجتمع وحدود الثقة :

كثيراً ما نحتاج إلى تقدير إجمالى الذين يتصفون بخاصية معينة باستخدام المعاينة الطبقية للنسب .

قمنا فيما سبق بتقدير نسبة المجتمع وحدى الثقة باستخدام الصيغتين (20 - 6) ، (5 - 6) . للحصول على تقدير القيمة الكلية ($\hat{T}$) نستخدم الصيغة التالية .

$$\hat{T} = N\, p_{st} \qquad \text{.... (6 - 23)}$$

أى تساوى :

$$\hat{T} = N \frac{1}{N} \sum_{h=1}^{L} N_h \, p_h$$

أى أن

$$\boxed{\hat{T} = \sum_{h=1}^{L} N_h \, p_h} \qquad \text{.... (6 - 24)}$$

أما الخطأ المعيارى لتقدير القيمة الكلية فيساوى :

$$V(\hat{T}) = N^2 \; V(p_{st})$$

ومقدره يساوى :

$$\hat{V}(\hat{T}) = N^2 \; \hat{V}(p_{st})$$

وبالتالى يكون حدا الثقة بمستوى ثقة % $(1 - \alpha)$:

$$\boxed{\hat{T} + Z_{(\alpha/2)} \sqrt{\hat{V}(\hat{T})} \leqslant T \leqslant \hat{T} + Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{T})}} \qquad \text{.... (6 - 25)}$$

تطبيق (٦-٢) :

أرادت الإدارة العليا فى إحدى المؤسسات دراسة مدى موافقة موظفيها حول الإجراءات الجديدة المتعلقة بالدوام . وقد تبين من نتائج الدراسة التى أجريت على عينة من الموظفين حجمها (١٠٠) موظف موزعين على مختلف مستويات الإدارة مايلى :

الطبقة	حجم المجتمع (N_h)	حجم العينة (n_h)	الموافقون
الإدارة العليا	١٠٠	١٠	٧
الإدارة المتوسطة	٣٠٠	٣٠	٢٤
الإدارة التنفيذية	٦٠٠	٦٠	٤٨
المجموع	١٠٠٠	١٠٠	٧٩

المطلوب :

١ – تقدير نسبة الموظفين الموافقين على الإجراءات الجديدة بمستوى ثقة ٩٥٪ .

٢ – تقدير إجمالى عدد الموظفين الموافقين على الإجراءات بمستوى ثقة ٩٥٪ .

الحـل :

– نقوم باستخراج تقدير نسبة المجتمع لكل طبقة ولإجمالى الطبقات :

أ – تقدير نسب الطبقات :

$$\hat{P}_h = p_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

$$p_h + q_h = 1$$

$$p_1 = \frac{7}{10} = 0.70 \qquad , \; q_1 = 1 - 0.70 = 0.30$$

$$p_2 = \frac{24}{30} = 0.80 \qquad , \; q_2 = 1 - 0.80 = 0.20$$

$$p_3 = \frac{48}{60} = 0.80 \qquad , \; q_3 = 1 - 0.80 = 0.20$$

ب – تقدير نسبة المجتمع

$$\hat{P} = p_{st} = \frac{\sum_{h=1}^{L} N_h \hat{P}_h}{N}$$

$$\hat{Q} = q_{st} = 1 - p_{st}$$

ويكون

$$p_{st} = \frac{(100 \times 0.70) + (300 \times 0.80) + (600 \times 0.80)}{1000}$$

$$= \frac{70 + 240 + 480}{1000} = \frac{790}{1000} = 0.79$$

وتكون

$$q_{st} = 1 - 0.79 = 0.21$$

جـ – حدا الثقة بمستوى ثقة (٩٥٪) :

$$p_{st} - ß \le P \le p_{st} + ß$$

$$ß = Z_{(1-\alpha/2)} \sqrt{\hat{V}(p_{st})}$$

$$\hat{V}(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{p_h q_h}{n_h}$$

$$= \frac{1}{(1000)^2} \left[\left(\frac{100 - 10}{99} \times 10000 \times \frac{0.70 \times 0.30}{10} \right) + \left(\frac{300 - 30}{299} \times 90000 \times \frac{0.80 \times 0.20}{30} \right) + \left(\frac{600 - 60}{599} \times 360000 \times \frac{0.80 \times 0.20}{60} \right) \right]$$

$$= \frac{1}{1000000} [190.9 + 433.4 + 865.4]$$

$$= \frac{1489.7}{1000000} = 0.0014897$$

$$ß = Z_{(1-\alpha/2)} \sqrt{\hat{V}(p_{st})}$$

$$= 1.96 \sqrt{0.0014897}$$

$$= 1.96 \times 0.038596$$

$$= 0.076$$

ويكون حدا الثقة :

$$0.79 - 0.076 \le P \le 0.79 + 0.076$$

$$0.714 \le P \le 0.866$$

أى أن تقدير نسبة الموظفين الموافقين على الإجراءات الجديدة بمستوى ثقة (٩٥٪) يتراوح بين (٧١,٤٪) و (٨٦,٦٪) من إجمالى الموظفين .

– لتقدير إجمالى عدد الموظفين الموافقين على الإجراءات نستخدم الصيغة التالية :

$$\hat{T} = N\, p_{st}$$

$$= 1000 \times 0.79 = 790$$

أى 790 موظفاً .

– حدا الثقة لإجمالى عدد الموظفين .

$$\hat{T} \pm Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{T})}$$

$$\hat{V}(\hat{T}) = N^2\, \hat{V}(p_{st})$$

$$= (1000)^2 \times 0.0014897$$

$$= 1489.7$$

ويكون حدا الثقة :

$$790 \pm 1.96\sqrt{1489.7}$$

$$790 \pm 76$$

أى أن

$$714 \le T \le 866$$

أى أن تقدير إجمالى الموظفين الموافقين على الإجراءات الجديدة للدوام يتراوح بين (٧١٤) موظفاً و (٨٦٦) موظفاً بمستوى ثقة (٩٥٪) .

ويمكن الحصول على الحدين نفسهما بضرب حدى الثقة للمتوسط بحجم المجتمع أى بـ (١٠٠٠) .

٦ – ٥ تحديد حجم العينة فى المعاينة الطبقية للنسب :

لتقدير نسبة المجتمع ، يجب أن نحدد المعلومات والبيانات التى نرغب فى الحصول عليها باستخدام المعاينة الطبقية ، إن صيغ حجم العينة (n) المناسب لتقدير نسبة المجتمع بخطأ

تقدير معين (β) هى الصيغ السابقة التى استخدمت عند تقدير متوسط المجتمع بعد تبديل σ_i^2 بما يوازيها فى النسب ($\sigma_i^2 = P_i Q_i$) .

إن حجم العينة اللازم لتقدير نسبة المجتمع إذا كان خطأ التقدير (β) هو

$$n = \frac{\sum_{h=1}^{L} \frac{N^2 P_h Q_h}{W_h}}{N^2 D + \sum_{h=1}^{L} N_h P_h Q_h} \quad \text{.... (6 - 26)}$$

حيث : $D = \frac{\beta^2}{Z^2}$ عندما نرغب فى تقدير نسبة المجتمع .

W_h كسر الوحدات المخصصة للطبقة (h) .

P_h نسبة المجتمع للطبقة (h) .

Q_h تساوى ($1-P_h$) أى نسبة الذين لا يتصفون بالخاصية للطبقة (h) .

أما الصيغة المستخدمة للتخصيص التى تجعل كلفة الوحدة أقل ما يمكن ·

$$n_h = \frac{N_h \sqrt{P_h Q_h / C_h}}{\sum_{h=1}^{L} N_h \sqrt{P_h Q_h / C_h}} \quad \text{.... (6 - 27)}$$

حيث (N_h) هو حجم الطبقة (h) فى المجتمع و (C_h) هى تكلفة الحصول على الوحدة فى الطبقة (h) .

ونستخدم تقديرات النسب (p_h) و (q_h) عندما تكون نسب المجتمع مجهولة فى الصيغ السابقة .

وتصبح الصيغ المستخدمة لتقدير نسبة المجتمع حسب طرق التخصص المختلفة كما يلى :

أ - حجم العينة باستخدام طريقة التخصيص المتناسب :

$$n = \frac{N \sum_{h=1}^{L} N_h P_h Q_h}{N^2 D + \sum_{h=1}^{L} N_h P_h Q_h} \quad .... (6 - 28)$$

ونستخدم تقديرات (P_h) و (Q_h) من بيانات العينة عندما تكون مجهولة أى نستخدم (p_h) و (q_h) . ويتم تخصيص حجم كل طبقة باستخدام الصيغة $n_h = n \frac{N_h}{N}$.

ب - حجم العينة باستخدام طريقة التخصيص المتساوى :

$$n = \frac{L \sum_{h=1}^{L} N_h^2 P_h Q_h}{N^2 D + \sum_{h=1}^{L} N_h P_h Q_h} \quad .... (6 - 29)$$

كذلك نستخدم (p_h) و (q_h) عندما تكون نسب المجتمع مجهولة . كما أن $n_h = \frac{n}{L}$.

جـ - حجم العينة استخدام طريقة التخصيص الأمثل :

$$n = \frac{\left[\sum_{h=1}^{L} N_h \sqrt{P_h Q_h C_h}\right]\left[\sum_{h=1}^{L} N_h \sqrt{P_h Q_h / C_h}\right]}{N^2 D + \sum_{h=1}^{L} N_h P_h Q_h} \quad .... (6 - 30)$$

ونستخدم (p_h) و (q_h) إذا كانت نسب المجتمع مجهولة ، ويتم تخصيص العينة باستخدام طريقة نيمان للتخصيص :

$$n = \frac{\left(\sum_{h=1}^{L} N_h \sqrt{P_h Q_h}\right)^2}{N^2 D + \sum_{h=1}^{L} N_h P_h Q_h} \qquad \text{.... (6 - 31)}$$

ونستخدم مقدرات نسب العينة إذا كانت نسب المجتمع مجهولة .

ويتم تخصيص حجم كل طبقة باستخدام الصيغة التالية إذا استخدمنا بيانات العينة :

$$n_h = n \frac{N_h \sqrt{p_h q_h}}{\sum N_h \sqrt{p_h q_h}} \qquad \text{.... (6 - 32)}$$

تطبيق (٦ - ٣) :

سحبت عينة استطلاعية لتقدير نسبة المدخنين فى أحدى الوزارات الموزعين حسب العمر وتبين مايلى :

	كبار العمر (٤٠ فاكثر)	الشباب (دون الـ٤٠ عامًا)	المجموع
عدد الموظفين	١٠٠	٢٠٠	٣٠٠
نسبة المدخنين (العينة الاستطلاعية)	٠,٤٠	٠,٦٥	١,٠
تكلفة الوحدة	٨	١٠	

المطلوب :

تحديد حجم العينة المناسب لتقدير نسبة المجتمع الذين يدخنون إذا كان خطأ التقدير المطلوب (٤٪) ، وذلك باستخدام طرق التخصيص التالية :

أ - المتناسب . ب - المتساوى .

جـ - الأمثل . د - نيمان .

وذلك بمستوى ثقة ٩٥ ٪ .

الحـــل :

أ - حجم العينة باستخدام طريقة التخصيص المتناسب :

$$n = \frac{N \sum_{h=1}^{L} N_h p_h q_h}{N^2 D + \sum_{h=1}^{L} N_h p_h q_h}$$

من بيانات التطبيق نجد أن .

$$D = \frac{\beta^2}{Z^2} = \frac{(0.04)^2}{(1.96)^2} = 0.000416$$

$$N_1 p_1 q_1 = 100 \times 0.40 \times 0.60 = 24$$

$$N_2 p_2 q_2 = 200 \times 0.65 \times 0.35 = 45.5$$

$$n = \frac{300 (24 + 45.5)}{(90000 \times 0.000416) + (24 + 45.5)}$$

$$n = \frac{300 \times 69.5}{37.44 + 69.5} = \frac{20850}{106.94}$$

$$= 195$$

ويكون حجم الطبقة الأولى :

لدينا

$$n_h = n \frac{N_h}{N}$$

$$n_1 = 195 \times \frac{100}{300} \approx 65$$

وحجم الطبقة الثانية :

$$n_2 = 195 \times \frac{200}{300} \approx 130$$

ب - حجم العينة باستخدام طريقة التخصيص المتساوى :

نستخدم الصيغة (29 - 6) بعد تبديل النسب بتقديراتها :

$$n = \frac{2\,[(10000 \times 0.4 \times 0.6) + (40000 \times 0.65 \times 0.35)]}{37.44 + 69.5}$$

$$= \frac{23000}{106.94} \approx 214$$

ويكون

$$n_1 = n_2 = \frac{214}{2} = 107$$

٣ - حجم العينة باستخدام طريقة التخصيص الأمثل :

نستخدم الصيغة (30 - 6) بعد تبديل النسب بتقديراتها :

الحد الأول من البسط يساوى (u_1) :

$$u_1 = [(100)\sqrt{0.4 \times 0.6 \times 8} + (200)\sqrt{0.65 \times 0.35 \times 10}]$$

$$= 138.56 + 301.66 = 440.22$$

والحد الثانى من البسط يساوى (u_2) :

$$u_2 = [(100)\sqrt{0.4 \times 0.6 / 8} + (200)\sqrt{0.65 \times 0.35 / 10})$$

$$= 17.32 + 30.17 = 47.49$$

ويكون حجم العينة :

$$n = \frac{440.22 \times 47.49}{106.94}$$

$$= \frac{20906.05}{106.94} = 195$$

وباستخدام الصيغة (29 - 6) نجد أن حجم الطبقات (باستخدام بيانات العينة) :

$$n_1 = 195 \times \frac{100\sqrt{0.4 \times 0.6/8}}{47.49} = 72$$

$$n_2 = 195 \times \frac{200\sqrt{0.65 \times 0.35 / 10}}{47.49} = 123$$

٤ – تحديد حجم العينة باستخدام طريقة نيمان للتخصيص :

حين نطبق الصيغة (31 - 6) باستخدام بيانات العينة ، نجد أن :

$$n = \frac{\left[100\sqrt{0.4 \times 0.6} + 200\sqrt{0.65 \times 0.35}\right]^2}{106.94} = 72$$

$$= \frac{(48.99 + 95.39)^2}{106.94} = 195$$

ويتم توزيع هذا الحجم على الطبقات باستخدام الصيغة (32 - 6) :

$$n_h = n \frac{N_h\sqrt{p_h q_h}}{\sum N_h \sqrt{p_h q_h}}$$

$$n_1 = \frac{195 \times 100\sqrt{0.4 \times 0.6}}{(100\sqrt{0.4 \times 0.6}) + (200\sqrt{0.65 \times 0.35})}$$

$$= \frac{195 \times 48.99}{48.99 + 95.39} = \frac{9553}{144.38} = 66$$

$$n_2 = \frac{195 \times 95.39}{144.38} = 129$$

الفصل
السابع

المعاينة المنتظمة

(Systematic Sampling)

٧-١ رموز وتعاريف :

نرغب أحيانًا فى تنفيذ بعض البحوث التى لا تتوافر عن مجتمعها بيانات دقيقة وشاملة كأسماء وعناوين الوحدات الإحصائية (الإطار) ، أو قد تتوافر فقط بيانات تقريبية عن حجم المجتمع . ويستخدم الإحصائيون فى مثل هذه الحالات ما يسمى المعاينة المنتظمة حيث نختار مثلاً واحداً من خمسة أو واحداً من عشرة . ولتوضيح هذا النوع من العينات نأخذ المثال التالى :

لدينا مجتمع مؤلف من (١٠٠) موظف ، نريد تقدير متوسط الدخل والإنفاق الشهرى للموظف باختيار عينة حجمها (١٠) موظفين . يوجد عدة طرق لاختيار وحدات هذه العينة إذ يمكن استخدام إحدى طرق الاختيار العشوائى كجداول الأرقام العشوائية مثلاً . ولكن يمكننا اختيار وحدات هذه العينة بالطريقة التالية المستخدمة عمليًا بشكل واسع : نختار عشوائيًا رقمًا يقع بين الصفر والعشرة ولنفرض أنه (٥) وذلك من وحدات المجتمع المائة المدرجة فى القائمة ، وبذلك تكون الوحدة الأولى فى العينة هى الموظف ذو الرقم (٥) ، أو بإضافة (١٠) إلى رقم الوحدة الأولى نحصل على رقم الوحدة الثانية وهو (١٥) ، وبإضافة (١٠) أيضًا يكون رقم الوحدة الثالثة (٢٥) وهكذا ... وتكون وحدات العينة المختارة :

٥ ، ١٥ ، ٢٥ ، ٣٥ ، ٤٥ ، ٥٥ ، ٦٥ ، ٧٥ ، ٨٥ ، ٩٥ .

إن العينة التى تم اختيارها بهذه الطريقة تسمى عينة منتظمة (واحد من ١٠) .

وقد يكون اختيار وحدات العينة المنتظمة حسب المكان أو الزمان أو الأحرف الأبجدية ، مثلاً ، لدراسة حالة طريق ما ، يمكننا تحديد نقطة ما على بعد (١٠ كلم) ثم نقوم بتحديد نقاط تبعد الواحدة عن الأخرى مسافة (٢٠) كلم وهكذا .....

وتستخدم المعاينة المنتظمة فى الحياة العملية بشكل واسع لقلة تكاليفها وسهولة اختيارها وقلة الأخطاء الناتجة عن اختيار وحدات العينة إذ تعد قليلة .

أما أهم عيوبها ، فهو عدم صلاحيتها إذا كانت الظاهرة المدروسة تتغير بصورة دورية لأن ذلك يعنى اختيار وحدات بشكل دورى ، والحصول على بيانات لا تمثل المجتمع فى هذه الحالة (إن مدى التأثير الدورى يعتمد على العلاقة بين طول الدورة وطول الفترة «K») . مثلاً ، لدراسة مساكن مبنية فى مجمعات سكنية ذات نمط متشابه ، فإن اختيار أحد المساكن فى المجموعة الأولى ، يعنى الحصول على مسكن مشابه فى المجموعة السكنية الثانية وهكذا ... وهذا يحد من الاستفادة من البيانات التى نحصل عليها باستخدام طريقة السحب المنتظم الموضحة سابقًا .

لقد استخدمت المعاينة المنتظمة فى بحوث كثيرة ، إذ ترفق أحياناً باستمارة التعداد العام للسكان استمارة تتضمن أسئلة للإجابة عليها من الأسر (مثلاً ١ من ٥) أى لكل خمس أسر سيجيب على استمارة العينة رئيس الأسرة .

كما يستخدم معهد غالوب لاستطلاعات الرأى العام هذا النوع من المعاينات فى بعض البحوث التى ينفذها .

ولتوضيح تعريف المعاينة المنتظمة نفترض لدينا مجتمع إحصائى مفرداته $X_1 , X_2 , \ldots , X_N$ وحجمه (N) وحدة مرتبة فى قائمة ما بشكلٍ ما (حسب المكان أو الزمان أو القيم) ونريد اختيار عينة حجمها (n) وحدة باستخدام المعاينة المنتظمة .

إذا رمزنا إلى طول الفترة بـ (K) وإلى ترتيب الوحدة الأولى المختارة عشوائيًا من الفترة الأولى بالرمز (j) فإن عدد العينات الممكن سحبها هو (K) عينة واحتمال سحب أى منها يساوى $\frac{1}{k}$.

تعرف المعاينة المنتظمة بأنها طريقة اختيار عدد من وحدات المجتمع عن طريق تقسيمه إلى (n) فترة (قسماً) وتحتوى كل فترة (K) وحدة بحيث يتم اختيار الوحدة الأولى عشوائيًا من الفترة الأولى ، وتتحدد أرقام الوحدات الأخرى للعينة على ضوء رقم الوحدة الأولى بإضافة (K) على رقم الوحدة المختارة وهكذا ... وذلك للاستدلال على خواص المجتمع كله عن طريق تعميم نتائج العينة .

٧ - ٢ طريقة اختيار العينة المنتظمة :

يمكننا تلخيص خطوات اختيار العينة المنتظمة بما يلى :

- تقسيم المجتمع إلى فترات (أقسام) عددها (n) فترة وحجم كل منها (K) وحدة وهكذا نجد أن :

$$K = \frac{N}{n} \qquad \text{أو} \qquad n = \frac{N}{K}$$

وتكون لدينا (n) فترة حجم كل منها (K) وحدة .

- نختار من وحدات (K) الأولى أى (K_1) وحدة باستخدام إحدى طرق الاختيار العشوائى (كجداول الأرقام العشوائية) ولنرمز إلى رقم (ترتيب) الوحدة الأولى المختارة بالرمز (j) .
- بعد اختيار الوحدة الأولى ذات الرقم (J) ، يتم تحديد أرقام الوحدات الأخرى للعينة المنتظمة وذلك بإضافة طول الفترة (K) إلى رقم الوحدة الأولى فنحصل على رقم

الوحدة الثانية $(J + K)$ ثم نضيف (K) إلى رقم الوحدة الثانية فيكون رقم الوحدة الثالثة $(J + 2K)$ وهكذا نكرر هذه العملية إلى أن نحصل على أرقام وحدات العينة المنتظمة وهي : $J , J + K , J + 2K , ... , J + (n - 1) K$

ويلاحظ أن ترتيب الوحدة الأولى في هذه العينة ، يحدد أرقام الوحدات الأخرى للعينة المنتظمة .

يشار أحيانًا إلى المعاينة المنتظمة بـ (١) من (K) (أى ١ من ٢٠ مثلاً) ويعنى ذلك أن طول الفترة (حجمها) هو K (أى عشرون مثلاً) حيث سنختار الوحدة الأولى من أرقام الوحدات العشرين الأولى ونضيف طول الفترة إلى كل ترتيب كما هو واضح فيما سبق .

ولتوضيح طريقة اختيار العينة المنتظمة ، نورد المثال التالى .

تطبيق (٧ – ١) :

تتكون إحدى الإدارات من (١٢) موظفًا ، سنوات خبراتهم كمايلى :

١ ، ٣ ، ٥ ، ٦ ، ٧ ، ٩ ، ١٢ ، ١٤ ، ١٥ ، ١٦ ، ١٨ ، ٢٢ نريد اختيار عينة منتظمة (١) من (٣) . ماهى مفردات العينة المنتظمة .

الحـــل :

إذا رمزنا لسنوات الخبرة بالرمز X يكون لدينا

$$X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 , X_8 , X_9 , X_{10} , X_{11} , X_{12}$$

نلاحظ من بيانات التطبيق أن $(K = 3)$ وبذلك يكون حجم العينة $n = \frac{N}{K} = \frac{12}{3} = 4$ أى أربع وحدات .

- نقسم وحدات المجتمع إلى أربع فترات طول كل منها ثلاث وحدات . الفترة الأولى هى (X_1 , X_2 , X_3) والفترة الثانية (X_4 , X_5 , X_6) والفترة الثالثة هى (X_7 , X_8 , X_9) والفترة الرابعة والأخيرة هـى $(X_{10} , X_{11} , X_{12})$.
- نختار من الفترة الأولى وحدة باستخدام إحدى طرق الاختيار العشوائى وستكون الوحدة المختارة إما الوحدة الأولى أو الوحدة الثانية أو الوحدة الثالثة ولتكن الوحدة المختارة هى الوحدة الثانية والتى قيمتها $(X_2 = 3)$.

– لتحديد ترتيب الوحدات الأخرى نضيف $(K = 3)$ إلى ترتيب الوحدة الأولى فيكون ترتيب الوحدة الثانية $(2 + 3)$ أي الوحدة الخامسة وقيمتها $(X_5 = 7)$ ثم نضيف إلى الرقم (5) طول الفترة (K) فتكون الوحدة الثالثة المختارة هي ذات الترتيب (8) وقيمتها $(X_8 = 14)$ ويكون ترتيب الوحدة الأخيرة $8 + 3 = 11$ وقيمتها $(X_{11} = 18)$ وتكون مفردات العينة المنتظمة : $x_1 = 3 , x_2 = 7 , x_3 = 14 , x_4 = 18$

ويمكننا القول إن عدد العينات الممكن سحبها تساوي (K) أي ثلاث عينات مفرداتها :

– العينة الأولى الممكن سحبها وتتكون من الوحدات ذات الترتيب (1, 4 , 7 , 10) ومفرداتها X_1 , X_4 , X_7 , X_{10} .

– العينة الثانية الممكن سحبها وتتكون من الوحدات ذات الترتيب (2, 5 , 8 , 11) ومفرداتها X_2 , X_5 , X_8 , X_{11} .

– العينة الثالثة الممكن سحبها وتتكون من الوحدات ذات الترتيب (3, 6 , 9 , 12) ومفرداتها X_3 , X_6 , X_9 , X_{12} .

إن احتمال سحب أية وحدة من الفترة الواحدة أي من (K) وحدة يساوي $\frac{1}{K}$ أي حسب التطبيق السابق $\frac{1}{3}$. ويلاحظ من التطبيق السابق أن حجم المجتمع هو من مضاعفات طول الفترة (K) أي $(N = nK)$ ولكن عملياً يوجد الكثير من الحالات التي لا يكون فيها حجم المجتمع من مضاعفات (K) أي أن $(N \neq nK)$. فإذا أردنا اختيار عينة من وحدات المجتمع واحد من خمسة نجد أن $(N = nK)$ وذلك لأن $n = \frac{12}{3} = 2.4$ وهذا يعني أن :

$$2 < n = \frac{12}{3} < 3$$

أي يتراوح حجم العينة بين وحدتين وثلاث وحدات وتكون مفردات العينات الممكن سحبها :

العينة الأولى	X_1 , X_6 , X_{11}
العينة الثانية	X_2 , X_7 , X_{12}
العينة الثالثة	X_3 , X_8
العينة الرابعة	X_4 , X_9
العينة الخامسة	X_5 , X_{10}

نلاحظ أن حجم بعض العينات هو ثلاث وحدات ، وحجم بعضها الآخر وحدتان فقط ونجد في حالة كون (N) ليست من مضاعفات (K) أن احتمال سحب أى وحدة يساوى $\frac{n}{N}$ وليس $\frac{1}{K}$ كما هو الحال فى كون (N) من مضاعفات (K) .

هناك طريقة أخرى تستخدم للتغلب على المشكلة التى تواجهنا عندما يكون حجم المجتمع ليس من مضاعفات (K) حيث نعتبر جميع الوحدات مرتبة على دائرة ، ونختار وحدة من (N) وحدة بشكل عشوائى ونضيف له طول الفترة إلى أن نحصل على الوحدات المختارة .

فى التطبيق السابق ذكرنا أن (K = 5 , N = 12) . نختار وحدة من وحدات المجتمع ولتكن الوحدة ذات الرقم (4) فتكون الوحدة الأولى ذات الترتيب (4) ونضيف (K = 5) فتكون ترتيب الوحدة الثانية (9 = 5 + 4) وترتيب الوحدة الثالثة (14 = 5 + 9) ، لكن (N = 12) لذا نترك الوحدة ذات الترتيب (1) ونأخذ الوحدة ذات الترتيب (2) وتكون الوحدات المختارة التى تمثل العينة المنتظمة المختارة هى الوحدات ذات الترتيب : 2 , 9 ,4 .

تستخدم المعاينة المنتظمة فى مجالات كثيرة كاستخدامها فى البحوث الاجتماعية التى تنفذ مع التعداد العام للسكان حيث يتم اختيار عينة من الأسر يتم تحديد أرقامها بشكل منتظم ، ثم ترفق استمارة العينة مع الاستمارة المخصصة للتعداد العام للسكان ، كذلك يمكن استخدامها لمراجعة الحسابات وأوامر الصرف أو فى مجال الخدمات والمرور وغيرها ، خاصة إذا كان حجم المجتمع غير معلوم بشكل دقيق .

٧ - ٢ تقديرات أهم معالم المجتمع :

٧ - ٢ - ١ تقدير متوسط المجتمع Estimation of Population Mean

سنقوم بتقدير متوسط المجتمع على أساس بيانات عينة منتظمة عندما (N = nK) . إن عدد العينات الممكن سحبها ، كما ذكرنا سابقًا هو (n) عينة واحتمال سحب أى منها يساوى (1/K) . إن مفردات العينة الممكن سحبها (i) حيث (i = 1 or 2 or . .. orK) هى $x_{i1} , x_{i2} , \dots\dots\dots , x_{in}$ ويكون الوسط الحسابى للعينة المنتظمة ولنرمز له بالرمز ($\bar{x}_{sy}$) يساوى :

$$\bar{x}_{sy} = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \qquad \dots (7-1)$$

j = 1 , 2 , 3 , ... , n
i = 1 or 2 or 3 or .... or K

إن متوسط العينة المنتظمة ($\bar{x}_{sy}$) هو مقدر غير متحيز لمتوسط المجتمع . وللتأكد من ذلك نعلم أنه يوجد لدينا (K) عينة ممكن سحبها ولكل منها متوسط أى لدينا :

$\bar{x}_1, \bar{x}_2, \ldots\ldots, \bar{x}_k$ واحتمال سحب أى منها يساوى 1/K . لذا نجد أن

$$E(\bar{x}_{sy}) = \frac{1}{K}(\bar{x}_1 + \bar{x}_2 + \ldots\ldots + \bar{x}_k)$$

$$= \frac{1}{K}\sum_{i=1}^{k}\bar{x}_i$$

$$= \frac{1}{K}\left(\frac{1}{n}\sum_{j=1}^{n}x_{1j} + \frac{1}{n}\sum_{j=1}^{n}x_{2j} + \ldots\ldots + \frac{1}{n}\sum_{j=1}^{n}x_{kj}\right)$$

$$= \frac{1}{K}\,\frac{1}{n}\,(x_1 + x_2 + \ldots\ldots + x_k)$$

حيث

$$x_i = \sum_{j=1}^{n} x_{ij}$$

$$= \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n} x_{ij} = \mu$$

٧ - ٣ - ٢ تقدير القيمة الكلية للمجتمع :

الصيغة المستخدمة لتقدير القيمة الكلية للمجتمع ولنرمز لها بالرمز $\hat{X}$ هى :

$$\boxed{\hat{X} = N\,\bar{x}_{sy}} \qquad \ldots (7-2)$$

$$= \frac{N}{n}\sum_{j=1}^{n} x_{ij}$$

حيث (i) ترمز إلى العينة الممكن سحبها ذات الترتيب (i) .

تطبيق (٧ - ٢) :

يبلغ عدد الحسابات المفتوحة في أحد البنوك (٥٠) حساباً . يرغب مدير البنك فى أخذ آراء أصحاب الحسابات حول مبالغ القروض التى يرغبون استلافها من البنك . وقد تم اختيار عينة منتظمة (واحد من عشرة كانت مفرداتها كما يلى بآلاف الريالات) .

١٠٠ ، ٣٠٠ ، ٢٠٠ ، ٤٠٠ ، ٥٠٠ .

المطلوب

تقدير متوسط مبلغ القرض الذى يرغب صاحب الحساب فى استلافه من البنك وتقدير إجمالى مبالغ هذه القروض .

الحـل

لدينا

$$\overline{x}_{sy} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

$$= \frac{1}{5}\,(100 + 300 + ... + 500)$$

$$= 300$$

أما تقدير إجمالى مبالغ القروض فيساوى :

$$\widehat{X} = N\ \overline{x}_{sy}$$

$$= 50 \times 300 = 15000$$

أى خمسة عشر مليون ريال

تطبيق (٧ - ٣)

لدينا تسعة أشخاص يملكون المبالغ التالية تم ترتيبهم تصاعدياً (بمئات الآلاف من الريالات) :
1,2,3,4,5,6,7,8,9

المطلوب :

١ – سحب عينة منتظمة واحد من ثلاثة وتوضيح العينات الممكن سحبها ومتوسطاتها .

٢ – إثبات أن تقدير متوسط المجتمع هو تقدير غير متحيز لمتوسط المجتمع .

الحـل :

العينات الممكن سحبها	المفردات	المتوسط $\bar{x}_{sy}$
العينة الأولى	1 , 4 , 7	4
العينة الثانية	2 , 5 , 8	5
العينة الثالثة	3 , 6 , 9	6

ونريد أن نثبت أن كلاً من هذه المتوسطات للعينات الممكنة هو مقدر غير متحيز لمتوسط المجتمع ، أى نريد إثبات أن

$$E(\bar{x}_{sy}) = \mu$$

إن احتمال سحب أية عينة من العينات الممكنة يساوى $\frac{k}{n}$ أى أن

$$P(\bar{x}_{sy}) = \frac{3}{9}$$

$$E(\bar{x}_{sy}) = \sum_{i=1}^{k} x_i \, p(\bar{x}_i)$$

$$E(\bar{x}_{sy}) = \left(\frac{3}{9} \times 4\right) + \left(\frac{3}{9} \times 5\right) + \left(\frac{3}{9} \times 6\right) = \frac{12}{9} + \frac{15}{9} + \frac{18}{9} = \frac{45}{9} = 5$$

ولحساب متوسط المجتمع ، نجد أن :

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1 + 25 + ... + 9}{9} = \frac{45}{9} = 5$$

أى أن

$$E(\bar{x}_{sy}) = \mu = 5$$

٧ – ٣ – ٢ تباين تقدير متوسط المجتمع :

يمكننا استخراج تباين تقدير متوسط المجتمع ولنرمز له بالرمز $V(\bar{x}_{sy})$ باستخدام عدة صيغ يتطلب معظمها معرفة جميع العينات الممكنة . لهذا السبب ، نجد أحياناً أن هناك صعوبة فى تقدير معلمات المجتمع من واقع بيانات عينة منتظمة . ونستعرض فيما يلى كيفية استخراج هذا التباين :

– من تعريف التباين يمكننا القول إن :

$$V(\bar{x}_{sy}) \ \frac{1}{k} \sum_{i=1}^{K} (\bar{x}_i - \mu)^2$$

حيث (K) عدد العينات الممكنة . ويمكن كتابة هذه الصيغة بالشكل الآتى ، كما يتضح فى الملحق رقم (٥ – ٣) فى نهاية الكتاب .

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \quad .... (7-3)$$

حيث

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \mu)^2$$

وهكذا نجد أن تباين مقدر متوسط المجتمع $V(\bar{x}_{sy})$ قد قسم إلى قسمين :

– تباين المجتمع $\frac{N-1}{N} S^2$

– التباين الداخلى وهو التباين بين الوحدات الواقعة داخل العينات المنتظمة الممكن سحبها . عندما يكون التباين الداخلى كبيراً ، فإن ذلك يعنى أن وحدات المعاينة فى (K) عينة منتظمة ، غير متجانسة . ولنوضح فيما يلى أثر التباين الداخلى على تباين تقدير متوسط المجتمع وعلاقته بتجانس أو عدم تجانس الوحدات . باستخدام معامل الارتباط داخل أزواج الوحدات الموجودة فى العينة المنتظمة الواحدة (r) والذى صيغته :

$$r = \frac{E(x_{ij} - \mu)(x_{ij'} - \mu)}{E(x_{ij} - \mu)^2}$$

حيث $(j < j')$

وباستخدام (r) نستطيع التعبير عن تباين تقدير متوسط المجتمع بالصيغة التالية ، كما يتضح من الملحق رقم (٥ –٥) :

$$V(\bar{x}_{sy}) = \frac{S^2}{n} \quad \frac{N-1}{N} [1 + (n-1) r] \quad .... (7-4)$$

حيث :

$$r = \frac{2}{(n-1)(N-1)S^2} \sum_{i=1}^{K} \sum_{j<j'}^{n} (x_{ij} - \mu)(x_{ij'} - \mu)$$

نلاحظ من الصيغة (4 - 7) أنه عندما يكون معامل الارتباط كبيراً أو موجباً فإن تباين تقدير متوسط المجتمع يكون كبيراً . وبالعكس عندما يكون معامل الارتباط صغيراً أو سالباً فإن التباين $V(\bar{x}_{sy})$ يكون صغيراً . أما عندما يكون معامل الارتباط مساوياً للصفر (r = 0) فإن تباين تقدير متوسط المجتمع للعينة المنتظمة $V(\bar{x}_{sy})$ يساوى تباين هذا التقدير للعينة العشوائية البسيطة $V(\bar{x}_{ran})$.

ويكون معامل الارتباط (r) كبيراً وموجباً عندما تكون الوحدات فى العينة المنتظمة متجانسة . ويكون هذا المعامل صغيراً وسالباً عندما تكون الوحدات فى العينة المنتظمة غير متجانسة .

– الصيغة الثانية المستخدمة لاستخراج تباين تقدير المتوسط للعينة المنتظمة هى :

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - K \frac{(n-1)}{N} S_w^2 \quad \text{.... (7 - 5)}$$

حيث (S^2) هو التباين المعدل للمجتمع و(S_w^2) التباين داخل وحدات العينات المنتظمة الممكنة ويساوى :

$$S_w^2 = \frac{1}{K(n-1)} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

ويلاحظ أن المقام فى صيغة التباين داخل وحدات العينات الممكنة يدل على وجود (K) عينة منتظمة كل منها يضيف (n - 1) درجة حرية لمجموع مربعات البسط . وللوصول إلى الصيغة (5 - 7) يمكن الرجوع إلى الملحق رقم (٥ – ٤) فى نهاية الكتاب .

تطبيق (٧ – ٤)

باستخدام بيانات التطبيق (٧ – ٣) ، استخرج تباين تقدير متوسط المجتمع ثم وضح درجة تجانس مفردات العينة المنتظمة .

الحـــل :

لدينا الصيغة التالية :

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

- استخراج قيمة الحد الأول من الطرف الأيسر :

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \mu)^2$$

$$= \frac{1}{9-1} [(1-5)^2 + (4-5)^2 + (7-5)^2 + ... + (6-5)^2 + (9-5)^2]$$

$$= \frac{1}{8} (21 + 18 + 21) = \frac{60}{8}$$

وبالتالي نجد أن الحد الأول يساوى

$$\frac{N-1}{N} S^2$$

$$= \frac{9-1}{9} \times \frac{60}{8} = \frac{20}{3}$$

أما الحد الثاني فيساوى :

$$\frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

$$= \frac{1}{9} [(1-4)^2 + (4-4)^2 + (7-4)^2 + (2-5)^2 + (5-5)^2$$

$$+ (8-5)^2 + (3-6)^2 + (6-6)^2 + (9-6)^2]$$

$$= \frac{1}{9} [(18 - 18 + 18) = \frac{54}{9} = 6$$

وبالتالى نجد أن تباين تقدير متوسط المجتمع يساوى :

$$V(\bar{x}_{sy}) = \frac{20}{3} - 6 = \frac{2}{3}$$

- يمكننا أيضاً حساب هذا التباين باستخدام الصيغة التالية :

$$V(\bar{x}_{sy}) = \frac{1}{K} \sum_{i=1}^{K} (\bar{x}_i - \mu)^2$$

$$= \frac{1}{3} \left[(4-5)^2 + (5-5)^2 + (6-5)^2 \right]$$

$$= \frac{1}{3} \times 2 = \frac{2}{3}$$

وهو الجواب السابق نفسه .

- استخراج تباين تقدير متوسط المجتمع باستخدام الصيغة

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{K(n-1)}{N} S_W^2$$

حيث

$$S_w^2 = \frac{1}{K(n-1)} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

أن

$$S_w^2 = \frac{1}{3(3-1)} \left[(1-4)^2 + (4-4)^2 + \ldots + (6-6)^2 + (9-6)^2 \right]$$

$$= \frac{1}{6} [54] = \frac{54}{6} = 9$$

وبالتالى يكون التباين داخل قيم وحدات العينات الممكن سحبها (الحد الثانى) :

$$\frac{K(n-1)}{N} S_w^2 = \frac{3(3-1)}{9} \times 9 = 6$$

ويكون تباين تقدير متوسط المجتمع

$$V(\bar{x}_{sy}) = \left(\frac{9-1}{9} \times \frac{60}{8}\right) - 6$$

$$= \frac{60-54}{9} = \frac{6}{9} = \frac{2}{3}$$

وهو الجواب السابق نفسه . ويلاحظ أن تباين المجتمع يساوى $\frac{60}{8}$ والتباين داخل وحدات العينات المنتظمة الممكن سحبها $\left(\frac{2}{3}\right)$ والفرق بينهما هو تباين تقدير متوسط المجتمع .

– لاستخراج تباين تقدير متوسط المجتمع لتوضيح درجة تجانس وحدات العينات المنتظمة نستخدم الصيغة التالية التى تحتوى على معامل الارتباط (r) :

$$V(\bar{x}_{sy}) = \frac{S^2}{n} \cdot \frac{N-1}{N}\left[1 + (n-1)\, r\right]$$

حيث :

$$r = \frac{2}{(n-1)(N-1)\, S^2} \sum_{i=1}^{K} \sum_{j<j'}^{n} (x_{ij} - \mu)\,(x_{ij'} - \mu)$$

لدينا

$$n = 3 \;,\; N = 9 \;,\; S^2 = \frac{60}{8}$$

بالنسبة للعينة الممكن سحبها الأولى نجد أن :

$$\sum_{j<j'} (x_{1j} - \mu)\,(x_{1j'} - \mu) = (x_{11} - \mu)\,(x_{12} - \mu) + (x_{11} - \mu)\,(x_{13} - \mu)$$

$$+ (x_{12} - \mu)\,(x_{13} - \mu)$$

$$= (1-5)\,(5-4) + (1-5)\,(7-5) + (4-5)\,(7-5)$$

$$= -6$$

كذلك نجد أن هذا المقدار مساوٍ للعينة الثانية (9 -) وللعينة الثالثة (6 -) ويكون معامل الارتباط

$$r = \frac{2}{(3-1)(9-1)(60/8)}(-6-9-6)$$

$$= \frac{-21 \times 8}{8 \times 60} = -21/60$$

إن معامل الارتباط سالب ومنخفض ، لذا تكون وحدات المعاينة غير متجانسة ويكون

$$V(\bar{x}_{sy}) = \frac{60/8}{3} \cdot \frac{9-1}{9}\left[1 + (3-1)\left(\frac{-21}{60}\right)\right]$$

$$= \frac{60 \times 8}{8 \times 3 \times 9}\left[1 - \frac{21}{30}\right] = \frac{20}{9} \times \frac{9}{30} = \frac{2}{3}$$

وهو الجواب السابق نفسه .

٧ - ٤ المقارنة بين المعاينة المنتظمة والمعاينات الأخرى وأشكال المجتمع :

٧ - ٤ - ١ المقارنة بين المعاينة المنتظمة والمعاينة العشوائية البسيطة :

لا بد لنا من توضيح مدى دقة المعاينة المنتظمة ومقارنتها مع الأنواع الأخرى للمعاينات كالمعاينة العشوائية البسيطة والمعاينة الطبقية .

يتضح من الصيغة التالية العلاقة بين تباين تقدير متوسط المعاينة المنتظمة وتباين المعاينة العشوائية البسيطة :

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2$$

ويمكننا إثبات أن* :

$$V(\bar{x}_{sy}) < V(\bar{x}_{ran})$$

إذا كان التباين داخل العينات المنتظمة الممكن سحبها أكبر من تباين المجتمع الكلي ، ويعني ذلك أن متوسط العينة المنتظمة هو أكثر دقة من متوسط العينة العشوائية البسيطة ،

* من أجل تفاصيل أكثر ، راجع : Cochran w. : Sampling Techniques , 1977 . (p. 208) .

وبالتالى تكون المعاينة المنتظمة أكثر دقة من المعاينة العشوائية البسيطة إذا كان التباين داخل العينات المنتظمة الممكن سحبها أكبر من تباين المجتمع كله .

ويمكن القول بشكل عام أن المعاينة المنتظمة تكون دقيقة عندما تكون الوحدات داخل العينة الواحدة غير متجانسة ، وتكون المعاينة المنتظمة غير دقيقة إذا كانت الوحدات متجانسة .

ويمكن القول كما يتضح من صيغة تباين تقدير متوسط المجتمع (4 - 7) ، إن العينة المنتظمة تكون ذات كفاية عظمى عندما يكون معامل الارتباط مساوياً (- 1) أى $(r = - 1)$. ويوجد لمعامل الارتباط أثرمهم على تباين العينة حتى لو كان بسيطاً فإن أثره يظهر بسبب العامل $(n - 1)$.

إن معرفة (r) فى المجتمع ليس بالأمر السهل ، لذا نجد صعوبة في مقارنة العينة المنتظمة بالعينة العشوائية البسيطة ، ولكن يمكن أحياناً افتراض قيمة معينة للارتباط ، وتجرى عملية المقارنة بين المعاينة العشوائية البسيطة والمعاينة المنتظمة . ويمكن القول إنه عندما يكون معامل الارتباط مساوياً للصفر ، فإن تباين تقدير متوسط العينة المنتظمة يساوى تقريباً تباين تقدير متوسط العينة العشوائية البسيطة . ويساعدنا هذا على استخدام تباين المعاينة العشوائية البسيطة كتقدير لتباين المعاينة المنتظمة ، وذلك لأننا وجدنا أنه لا يمكن إيجاد مقدر غير متحيز لتباين متوسط العينة المنتظمة من بيانات عينة منتظمة واحدة .

أما كفاءة المعاينة المنتظمة بالنسبة للمعاينة العشوائية البسيطة فتتضح من :

$$\frac{V(\bar{x}_{sy})}{V(\bar{x}_{ran})} = \frac{(N - 1)[1 + (n - 1) r]}{n (K - 1)} \quad .... (7 - 6)$$

وعندما يكون الناتج الواحد الصحيح ، فهذا يعنى أن دقة العينتين متشابهتان ، وفى هذه الحالة نجد من الصيغة السابقة أن $r = \frac{-1}{N - 1}$ وهذا يعنى عندما يساوى معامل الارتباط (r) المقدار $\left(\frac{-1}{N - 1}\right)$ فإن استخدام المعاينة المنتظمة والمعاينة العشوائية البسيطة سيعطى الدقة نفسها . لذا عندما يكون معامل الارتباط صغيراً فى حالة كون (N) كبيرة ، فإننا نستخدم $V(\bar{x}_{ran})$ عوضاً عن $V(\bar{x}_{sy})$ وهذه نتيجة مهمة إذ ستمكننا من تقدير تباين تقدير متوسط العينة المنتظمة باستخدام :

$$V(\bar{x}_{ran}) = \frac{(N - 1)}{N} \frac{s^2}{n} \quad .... (7 - 7)$$

حيث (s^2) هو تباين العينة ويساوى :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

وعمليًا نقوم بحساب تباين العينة ونعوضه فى الصيغة (7 - 7) للحصول على تقدير تباين العينة المنتظمة أى $\hat{V}(\bar{x}_{sy})$. ويمكننا إثبات أن تباين العينة العشوائية البسيطة فى هذه الحالة هو مقدر غير متحيز لتباين العينة المنتظمة $\hat{V}(\bar{x}_{sy})$ كما يتضح من ملحق هذا الكتاب (٥ – ٣) .

٧ – ٤ – ٢ العلاقة بين المعاينة المنتظمة والمعاينة الطبقية العشوائية :

تعد المعاينة المنتظمة حالة خاصة من المعاينة الطبقية العشوائية ، ويمكننا القول إن المجتمع الذى نقوم بدراسته قد قسم إلى طبقات عددها (n) طبقة ، وتحتوى كل طبقة (K) من المفردات بحيث يتم اختيار وحدة واحدة من كل طبقة عوضًا عن (n_h) أى حجم العينة للطبقة (h) ويتم فيها ـ كحالة خاصة ـ اختيار وحدة واحدة عشوائيًا من الطبقة الأولى وعلى ضوئها تتحدد أرقام الوحدت المختارة ونحصل على عينة منتظمة . أما إذا اخترنا وحدة واحدة عشوائيًا من كل طبقة ، فالعينة التى نحصل عليها هى عينة طبقية . لذا فإن المعاينة المنتظمة هى حالة خاصة من المعاينة الطبقية العشوائية يتم فيها الاختيار العشوائى من الطبقة الأولى فقط لوحدة واحدة وتتحدد أرقام الوحدات المختارة على ضوء رقم الوحدة المختارة .

٧ – ٤ – ٢ المعاينة المنتظمة وأشكال المجتمع * :

تختلف دقة المعاينة المنتظمة من مجتمع لآخر ، إذ تعد هذه المعاينة ذات دقة عالية فى بعض المجتمعات وتعد ذات دقة منخفضة فى بعضها الآخر حيث يفضل استخدام أنواع أخرى من المعاينات كالمعاينة العشوائية البسيطة أو المعاينة الطبقية العشوائية أو غيرهما .

لذا لابد من التعرف على تركيب وطبيعة المجتمع الذى ندرسه ، ونستطيع التمييز بين أربعة أشكال من المجتمعات :

---

* للحصول على تفاصيل أكثر ، راجع : Cochran W. : Sampling Techniques , pp 212 - 221

أ – المجتمعات ذات الترتيب العشوائى :

يقصد بالمجتمعات المرتبة عشوائيًا المجتمعات المدرجة فى الإطار بشكل لا يوجد علاقة بين قيم مفردات المجتمع وقائمة أسمائها المدونة عشوائيًا . ويلاحظ فى هذه المجتمعات عدم وجود علاقة بين الخاصية المقاسة وتنظيم وحدات المجتمع ، كما لا يوجد ارتباط بين الوحدات المتجاورة . نجد فى هذه الحالة أن المعاينة المنتظمة تكون مكافئة للمعاينة العشوائية البسيطة ، إذ ستكون وحدات هذه العينة غير متجانسة وسيكون معامل الارتباط فيها صغيراً . وعندما يكون هذا المعامل صغيراً ، فإن تباين المعاينة العشوائية البسيطة وتباين المعاينة المنتظمة سيكونان متساويين (أو على الأقل يتساويان فى المتوسط) .

ب – المجتمع المرتب (المنظم) :

عندما يتم ترتيب الوحدات حسب الخاصية المدروسة ونسحب عينة منتظمة ، نحصل على وحدات غير متجانسة . إن المجتمع الذى نختار منه هذه العينة هو مجتمع مرتب . مثلاً ، إذا رتبنا الحيازات الزراعية حسب المساحة يكون لدينا مجتمع مرتب أو منظم . ويلاحظ وجود علاقة بين الخاصية المقاسة وقائمة أسمائها (الإطار) . إن تباين المعاينة المنتظمة المختارة من مجتمعات مرتبة سيكون أصغر من تباين المعاينة العشوائية البسيطة ، أى أن وحدات المعاينة المنتظمة المسحوبة من مجتمع منظم ستكون أقل تجانسًا من وحدات المعاينة العشوائية البسيطة المسحوبة من المجتمع نفسه وهذا يعنى أن معامل الارتباط سيكون صغيراً . ويمكننا القول إن : $V(\bar{x}_{sy}) < V(\bar{x}_{ran})$.

ج – المجتمعات ذات الاتجاه الخطى :

إذا كانت قيمة وحدات المجتمع ذات اتجاه خطى حيث تزيد أو تنقص كل وحدة عن الوحدة التى تسبقها بمقدار ثابت (تقريبًا) فإننا نجد أن المعاينة المنتظمة أفضل من المعاينة العشوائية البسيطة ، كما أن المعاينة الطبقية أفضل من المعاينة المنتظمة أى أن :

$$V(\bar{x}_{st}) \le V(\bar{x}_{sy}) \le V(\bar{x}_{ran})$$

وذلك لأنه إذا كان يوجد فى العينة المنتظمة قيم منخفضة فى إحدى الطبقات ، فإن قيمها فى الطبقات الأخرى تكون أيضًا منخفضة ، بينما تعطى المعاينة الطبقية الفرصة للأخطاء داخل الطبقة الواحدة لتحذف بعضها البعض . ونستطيع إزالة أثر الاتجاه الخطى فى حالة استخدام المعاينة المنتظمة باختيار قيمة مركزية لترتيب المفردة بدلاً من اختيار هذه القيمة

عشوائيًا . كما أن هناك طريقة أخرى تعرف بطريقة تصحيح النهايات التى يتم بموجبها تبديل المتوسط غير المرجح بمتوسط مرجح بالمقدار $\left(\frac{1}{n}\right)$ ماعدا القيمتين الأولى والأخيرة اللتين تأخذان أوزانًا أخرى . وتسمى هذه الأوزان (الترجيحات) «أوزان تصحيح النهايات» ومن هذه التصحيحات «تصحيح ياتس Yates» الذى يستخدم أوزانًا تختلف عن تلك الموضحة فى الطريقة السابقة .

د – المجتمعات ذات التغيرات الدورية :

نجد فى بعض الأحيان ، أن وحدات المعاينة فى المجتمع ذات اتجاه دورى وأثر وحدات المعاينة المختارة يتوقف فى هذه الحالة على قيمة (K) أى طول الفترة . نجد فى هذه الحالة أن قيم وحدات العينة المنتظمة متشابهة ومتجانسة ويكون معامل الارتباط (r) كبيراً . مثلاً عندما يكون لدينا ثلاث فترات مفرداتها :

1, 2, 3, 4   1, 2, 3, 4   1, 2, 3, 4

عند اختيار المفردة الثالثة نجد أن مفردات العينة المنتظمة هى (3,3,3) وهى متشابهة . لحل هذه المشكلة ، لابد من تغيير مكان وحدة المعاينة بشكل يمكننا من الحصول على مفردات غير متشابهة . مثلاً إذا كان ترتيب المفردة المختارة الأولى (j) فإننا نضيف (k + 1) عوضاً عن (K) فيكون ترتيب المفردات هكذا :

$$j, j + K + 1, j + 2k + 2, \ldots, j + (n - 1) K + (n - 1)$$

٧ – ٤ – ٤ تقدير تباين تقدير متوسط المجتمع :

إن استخراج تباين تقدير متوسط المجتمع باستخدام الصيغ السابقة مستحيل فى معظم الأحيان ، خاصة إذا كان حجم المجتمع كبيراً إذ لا يمكن معرفة العينات الممكن سحبها . لذا لابد لنا من تقدير هذا التباين من بيانات عينة منتظمة يتم اختيار وحداتها من بيانات المجتمع ، ويمكننا استخدام الصيغ التالية لتقدير تباين تقدير متوسط المجتمع ، ولنرمز له بالرمز $\hat{V}(\overline{x}_{sy})$:

١ – إذا كانت المجتمعات ذات الترتيب العشوائى ، يمكننا استخدام تقدير تباين تقدير متوسط المجتمع الذى صيغته :

$$\hat{V}(\overline{x}_{sy}) = \frac{N - n}{Nn} \times \frac{\sum_{i=1}^{n} (x_i - \overline{x}_{sy})^2}{n - 1} \quad \ldots (7 - 8)$$

ويعد هذا المقدر مقدراً غير متحيز لتباين تقدير متوسط المجتمع . وهذه الصيغة هي الصيغة نفسها المستخدمة للعينة العشوائية البسيطة .

٢ – في المجتمعات ذات الاتجاه الخطي ، نستخدم الصيغة التالية :

$$\hat{V}(\bar{x}_{sy}) = \frac{N - n}{N} \times \frac{n'}{n^2} \times \frac{\sum (x_i - 2x_{i+k} + x_{i+k2})^2}{6(n-2)} \quad .... (7-9)$$

حيث

$$(1 \leq i \leq n-2)$$

أن n'/n^2 هي مجموع مربعات الأوزان في المتوسط المرجح (تصحيح النهايات) وإذا كانت (n) كبيرة ، يمكن استبدال هذا المجموع بالعامل $\frac{1}{n}$ لان الفترة (الطبقة) في النهايات لها وزن ترجيحي صغير ويكون التقدير متحيزاً ما لم يكن (σ^2_i) ثابتاً ، ولكنه يعد مقبولاً إذا كان (n) كبيراً والنموذج خطي .

٣ – لقد اقترح ياتس (Yates) في عام ١٩٤٩ [*] مقدراً يعتمد على الفروق (d_u) ، إذ قام بتقسيم العينة المتتابعة إلى مجموعات تشمل كل منها (٩) مفردات ، الأولى من (١) إلى (٩) والثانية (٩) إلى (١٧) ، واستخدم الترجيح (الوزن) لكل من المفردة الأولى والأخيرة . وقام بإعداد الفروق ... , d_1, d_2 حيث

$$d_1 = (\frac{1}{2}x_1 + x_3 + x_5 + x_7 + \frac{1}{2}x_9) - (x_2 + x_4 + x_6 + x_8)$$

وتبدأ d_2 بـ x_9 ونتبع الأسلوب نفسه لحساب بقية الفروق . ونستخدم الصيغة التالية لاستخراج تقدير تباين تقدير متوسط المجتمع :

$$\hat{V}(\bar{x}_{sy}) = \frac{N - n}{Nn} \cdot \frac{\sum_{u=1}^{g} d_u^2}{7.5g} \quad .... (7-10)$$

* من أجل تفاصيل أكثر ، راجع : Cochran W. , Sampling Techniques , P. 226 .

حيث (7.5) هو مجموع مربعات المعاملات فى أى من الفروق (d_u) و(g) هو عدد الفــروق الموجودة فى العينة ($g = n$) .

إن الطرق السابقة المستخدمة لتقدير تباين تقدير متوسط المجتمع من بيانات عينة منتظمة تعطى تقديرات غير دقيقة للتباين . لذا نجد أن البعض يفضل استخدام العينة المنتظمة مع الأنواع الأخرى للعينات .

٤ – يمكننا تقدير تباين تقدير متوسط المجتمع باستخدام عدد من العينات المنتظمة المتكررة كما سيتضح فى نهاية هذا الفصل عند شرح هذا النوع من العينات .

٧ – ٥ حدود الثقة لتقديرات متوسط المجتمع والقيمة الكلية للمجتمع :

للاستفادة من بيانات العينة المنتظمة بشكل أفضل ، لابد من تقدير حدى الثقة وذلك بمستوى ثقة معين % ($1 - \alpha$) .

وتستخدم الصيغة التالية ، لاستخراج حدى الثقة لتقدير متوسط المجتمع :

$$\bar{x}_{sy} \pm t_{1-\alpha/2, n-1} \sqrt{\hat{V}(\bar{x}_{sy})} \qquad .... (7 - 11)$$

حيث (t) تمثل القيم المستخرجة من جداول توزيع (t) عندما يكون حجم العينة صغيراً (أقل من ٣٠) وذلك بمستوى ثقة % ($1 - \alpha$) ودرجــات حرية ($n - 1$) أما عندما يكون حجم العينة (٣٠) فأكثر ، نستخدم القيم المستخرجة من جداول التوزيع الطبيعى (Z) . أما $\hat{V}(\bar{x}_{sy})$ فنستخدم إحدى الصيغ الموضحة فيما سبق حسب نوع المجتمع .

كذلك نجد أن حدّى الثقة لتقدير القيمة الكلية للمجتمع يساويان :

$$\hat{T} + Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{T})} \qquad .... (7 - 12)$$

حيث

$$\hat{T} = N \bar{x}_{sy}$$

$$\hat{V}(\hat{T}) = \hat{V}(N \bar{x}_{sy})$$

$$= N^2 \hat{V}(\bar{x}_{sy})$$

و Z هى القيمة المستخرجة من جداول التوزيع الطبيعى بمستوى ثقة معين . ولاستخراج $\hat{V}(\overline{x}_{sy})$ نستخدم إحدى الصيغ السابقة حسب شكل المجتمع الذى نقوم باختيار العينة منه .

تطبيق (٧-٥) :

استخرجت البيانات التالية من نتائج عينة منتظمة (١) من (٢٠) وذلك لتقدير متوسط مدة التدريب التى قضاها الموظفون الملتحقون بدورات تدريبية خلال العام الماضى (بالأشهر) فى إحدى الوزارات :

$n = 10 , N = 200 , \overline{x} = 5 , s^2 = 4$

المطلوب استخراج حدى الثقة لمتوسط مدة التدريب وإجمالى سنوات التدريب .

الحل

نستخدم الصيغة (7 - 7) :

$$\hat{V}(\overline{x}_{sy}) = \frac{N - n}{N} \frac{s^2}{n}$$

$$= \frac{200 - 10}{200} \frac{4}{10}$$

$$= \frac{760}{2000} = 0.38$$

ويكون حد الثقة :

$$\overline{x}_{sy} \mp t_{1-\alpha/2, n-1} \sqrt{\hat{V}(\overline{x}_{sy})}$$

$$5 \mp 2.262 \sqrt{0.38}$$

$$5 \mp 1.394$$

أى أن متوسط شهور التدريب التى قضاها الموظفون تقع بين (٣،٦٠٦) و (٦،٣٩٤) أشهر وذلك بمستوى ثقة ٩٥٪ أى :

$$3.606 \le \mu \le 6.394$$

– لتقدير إجمالى سنوات التدريب نعلم أن

$$\hat{V}(\hat{T}) = \hat{V}(N\,\bar{x}_{sy})$$

$$= N^2\,\hat{V}(\bar{x}_{sy})$$

$$= (200)^2\,(0.38) = 15200$$

$$\hat{T} = N\,\bar{x}_{sy}$$

$$= 200 \times 5 = 1000$$

ويكون حدا الثقة

$$\hat{T} \mp Z_{1-\alpha/2}\sqrt{\hat{V}(\hat{T})}$$

$$1000 \mp 1.96\sqrt{15200}$$

$$1000 \mp 241.64$$

أى أن إجمالى شهور التدريب التى قضاها الموظفون تقع بين (٧٥٨,٣٦) و (١٢٤١,٦٤) شهراً وذلك بمستوى ثقة ٩٥٪ أى .

$$758.36 \leqslant T \leqslant 1241.64$$

٧ – ٦ **تقدير نسبة المجتمع :** (Estimation of A Population Proportion)

كثيراً ما يرغب الباحث فى استخدام بيانات عينة منتظمة لتقدير نسبة أفراد المجتمع الذين يتصفون بخاصية معينة . قد ترغب ـ مثلاً ـ فى تقدير نسبة الذين يؤيدون قراراً معيناً خاصة إذا كان حجم المجتمع غير محدد بشكل دقيق . نختار فى هذه الحالة والحالات المشابهة عينة منتظمة (١) من (K) من قوائم الناخبين (إذا كانت متوافرة) .

إذا رمزنا لنسبة المجتمع بالرمز (P) ولتقدير هذه النسبة من بيانات عينة منتظمة بالرمز ($\hat{P}_{sy}$) فإن

$$\hat{P}_{sy} = \bar{x}_{sy} = \frac{1}{n}\sum_{i=1}^{n} x \qquad \text{.... (7 - 13)}$$

حيث

$x_i = 1$ إذا كانت الوحدة تتصف بالخاصية التى ندرسها .

$x_i = 0$ إذا كانت الوحدة لا تتصف بالخاصية التى ندرسها .

ويكون تقدير نسبة الذين لا يتصفون بالخاصية $(\hat{Q}_{sy})$:

$$\hat{Q}_{sy} = 1 - \hat{P}_{sy} \qquad .... (7 - 14)$$

أما تقدير تباين تقدير نسبة المجتمع فيساوى :

$$\hat{V}(\hat{P}_{SY}) \frac{\hat{P}_{sy} \quad \hat{Q}_{sy}}{n - 1} \left(\frac{N - n}{N}\right) \qquad .... (7 - 15)$$

ويكون حدا الثقة بمستوى ثقة % $(1-\alpha)$:

$$\hat{P}_{sy} \mp Z_{1-\alpha/2} \sqrt{\hat{V}(\hat{P}_{sy})} \qquad .... (7 - 16)$$

ويمكننا تجاهل معامل تصحيح المجتمع المحدود $(N - n)/N$ إذا كان حجم المجتمع غير معلوم وكبيراً بالنسبة لحجم العينة (n) .

تطبيق (٧ – ٦) :

ترغب إحدى المؤسسات فى تقدير نسبة زبائنها الذين يؤيدون زيادة التسهيلات المالية التى تمنح لهم . وقد بلغ عدد المؤيدين لهذا الاقتراح (٢٠٠) شخص من بين عينة حجمها (٣٠٠) زبون علماً بأن عدد الزبائن هو (٤٠٠٠) زبون .

ماهو تقدير نسبة الزبائن الذين يؤيدون زيادة التسهيلات المالية وما هو عددهم المقدر ؟

الحــل

$$\hat{P}_{sy} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{300}(200) = 0.667$$

$$\hat{Q}_{sy} = 1 - \hat{P}_{sy}$$

$$= 1 - 0.667 = 0.333$$

$$\hat{V}(\hat{P}_{sy}) = \frac{0.667 \times 0.333}{200 - 1} \frac{(4000 - 200)}{4000}$$

$$= \frac{844.02}{796000} = 0.00106$$

ويكون حدا الثقة للنسبة :

$$\hat{P}_{sy} \mp Z_{1-\alpha/2} \sqrt{\hat{V}(\hat{P}_{sy})}$$

$$0.667 \mp 1.96 \sqrt{0.00106}$$

$$0.667 \mp 0.0638$$

أى بدرجة ثقة ٩٥٪ ، نجد أن نسبة المؤيدين لزيادة التسهيلات المالية تتراوح بين 0.6032 و 0.7308 .

$$0.6032 \leq P \leq 0.7308$$

أما تقدير عدد الزبائن المؤيدين لهذه التسهيلات :

$$\hat{T} = N \hat{P}_{sy}$$

$$= 4000 \times 0.667$$
$$= 2668$$

ويكون

$$\hat{V}(\hat{T}) = N^2 \hat{V}(\hat{P}_{sy})$$

$$= 4000^2 \times 0.00106$$

$$= 16960$$

ويكون حدا الثقة

$$\hat{T} \mp Z_{1-\alpha/2} \sqrt{\hat{V}(\hat{T})}$$

$$266.8 \mp 1.96 \sqrt{16960}$$

$$2668 \mp 255$$

أى أن تقدير إجمالى المؤيدين بمستوى ثقة ٩٥٪ يتراوح بين 2413 و 2923 أى أن

$$2413 \leq T \leq 2923$$

ويمكننا الحصول على النتائج نفسها بضرب حدى الثقة لنسبة المؤيدين للتسهيلات بحجم المجتمع أى :

$$4000 \times 0.6032 \leq T \leq 4000 \times 0.7308$$

$$2413 \leq T \leq 2923$$

٧ - ٧ تحديد حجم العينة المنتظمة :

نستخدم الصيغة التالية لاستخراج حجم العينة المنتظمة المطلوب لتقدير متوسط المجتمع :

$$\beta = Z_{(1-\alpha/2)} \sqrt{V(\bar{\bar{x}}_{sy})}$$

حيث (β) هو حد خطأ التقدير . ولحل هذه المعادلة ، يجب معرفة التباين (σ^2) ومعامل الارتباط (r) أو قيمة تقريبية لهما . ويمكننا استخدام تباين العينة الاستطلاعية لتقدير تباين المجتمع ، ونجد أن حجم العينة (n) يساوى :

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad \text{.... (7 - 17)}$$

حيث $D = \frac{\beta^2}{Z^2}$. ونستخدم الصيغة نفسها لتحديد حجم العينة اللازم لتقدير

القيمة الكلية للمجتمع ، ولكن تصبح D فى هذه الحالة $D = \frac{B^2}{Z^2 N^2}$ كذلك فى حال النسب نضع PQ عوضاً عن σ^2 فى الصيغة السابقة .

تطبيق (٧ – ٧) :

تبين من دراسة سابقة أن متوسط رواتب الموظفين فى إحدى الوزارات هو (٥٠٠٠) ريال والانحراف المعيارى هو (٢٠٠) ريال . ما هو حجم العينة المنتظمة المطلوب لتقدير متوسط المجتمع إذا كان حد خطأ التقدير المحدد (١٠٠) ريال وعدد موظفى الوزارة (٥٠٠٠) موظف (بمستوى ثقة ــ ٩٥٪) .

الحــل :

لدينا

$$\mu = 5000 \ , \sigma = 200 \ , B = 100 \ , N = 5000$$

لتحديد حجم العينة نستخدم الصيغة

$$n = \frac{N \sigma^2}{(N - 1) D + \sigma^2}$$

$$D = \frac{B^2}{Z^2} = \frac{(100)^2}{(1.96)^2} = \frac{10000}{3.8416}$$

$$= 260.31$$

$$n = \frac{5000 \times (200)^2}{(5000 -) (260.31) + (200)^2}$$

$$= \frac{200000000}{1341290}$$

$$= 149.11 \approx 149$$

أى أن حجم العينة اللازم لتقدير متوسط المجتمع هو (١٤٩) موظفاً .

تطبيق (٧ – ٨) :

تبين من عينة استطلاعية لتقدير نسبة المدخنين فى إحدى الكليات أن (٣٠) طالباً يدخنون من بين (٧٥) طالباً . ما هو حجم العينة المناسب لتقدير نسبة المدخنين فى الكلية إذا كان حد خطأ التقدير المحدد (٣٪) وبمستوى ثقة (٩٥٪) إذا كان عدد طلاب الكلية (٥٠٠٠) طالب .

الحــل : لدينا

$$\hat{P} = \frac{30}{75} = 0.40$$

$$\hat{Q} = 1 - \hat{P}$$

$$= 1 - 0.40 = 0.60$$

$$n = \frac{N \hat{P} \hat{Q}}{(N - 1) D + \hat{P} \hat{Q}}$$

$$D = \frac{B^2}{Z^2} = \frac{(0.03)^2}{(1.96)^2} = \frac{0.0009}{3.8416} = 0.00023$$

$$n = \frac{5000 \times 0.40 \times 0.60}{(5000 - 1) \times (0.00023 + (0.60 \times 0.40)}$$

$$= \frac{1200}{1.38977}$$

$$= 863.45 = 863$$

ويكون حجم العينة النهائى

$$n = \frac{no}{1 + \frac{no}{N}}$$

$$= \frac{863}{1 + \frac{863}{5000}} = \frac{863}{1.1726} \approx 736$$

٧ - ٨ المعاينة الطبقية المنتظمة : (Stratified Systematic Sampling)

رأينا فيما سبق أن ترتيب الوحدات بشكل مناسب يعطي نوعًا من التصنيف الطبقي بكسر معاينة متساوٍ . إذا قسمنا المجتمع إلى طبقات وفقًا لمعايير أخرى ، فإننا قد نختار عينة منتظمة منفصلة لكل طبقة بعد تحديد ترتيب الوحدة الأولى في كل طبقة ، ويسمى هذا النوع من العينات «المعاينة الطبقية المنتظمة» . وهذه العينة مناسبة إذا أردنا الحصول على تقديرات لكل طبقة أو إذا استخدمت كسور معاينة غير متساوية . وتعد هذه الطريقة أكثر دقة من العينة الطبقية العشوائية البسيطة إذ المعاينة المنتظمة داخل كل طبقة هي أكثر دقة من العينة العشوائية البسيطة الموجودة داخل الطبقة .

إذا رمزنا لمتوسط العينة الطبقية المنتظمة بالرمز ($\overline{x}_{stsy}$) فإن مقدر متوسط المجتمع ومقدر تباينه يساويان :

$$\overline{x}_{stsy} = \frac{\sum_{h=1}^{L} N_h \overline{x}_{syh}}{N}$$

$$V(\overline{x}_{stsy}) = \frac{\sum_{h=1}^{L} N_h^2 V(\overline{x}_{syh})}{N^2} \qquad .... (7 - 18)$$

حيث (N_h) حجم الطبقة (h) في المجتمع و ($\overline{x}_{syh}$) يساوي متوسط العينة المنتظمة للطبقة (h) و ($V(\overline{x}_{syh})$) هو تباين تقدير متوسط المجتمع للطبقة (h) . وتستخدم إحدى الصيغ المستخدمة لتقدير هذا التباين الموضحة فيما سبق .

٧ - ٩ المعاينة المنتظمة المتكررة : (Repeated Systematic Sampling)

ذكرنا فيما سبق أنه من الممكن استخراج تباين تقدير متوسط المجتمع ($V(\overline{x}_{sy})$) من بيانات عينة منتظمة واحدة في حالة اعتبارها كعينة عشوائية بسيطة إذا كانت $(r) = \frac{-1}{N-1}$ التي تكون فيها العينة المنتظمة مكافئة للعينة العشوائية البسيطة . ولكن في كثير من الحالات ، نجد أن المعاينة المنتظمة ليست كفؤًا للمعاينة العشوائية البسيطة ، لذا نجد أن هناك طريقة أخرى لتقدير التباين باستخدام ما يسمى المعاينة المنتظمة المتكررة .

كما يتضح من اسم هذه المعاينة ، يتطلب هذا النوع من المعاينات اختيار أكثر من معاينة منتظمة بحيث يكون مجموع أحجامها يساوى حجم العينة المنتظمة التى نريد استبدالها .

ولتوضيح هذا النوع من المعاينات نفترض أننا نرغب فى اختيار عينة حجمها (٨٠) من مجتمع يتضمن (٤٠٠) وحدة معاينة ($N = 400$) . فى هذه الحالة نجد أن ($K = \frac{N}{n} = \frac{400}{80} = 5$) ، أى نرغب اختيار عينة منتظمة واحد من خمسة . ولكن يمكننا اختيار أكثر من عينة واحدة (ثلاث أو خمس أو ثمانى) عينات متساوية فى الحجم ، ومجموع أحجامها يساوى حجم العينة أى ($n = 80$) . مثلاً إذا قررنا اختيار (8) عينات حجم كل منها ($n' = 10$) ويكون $K' = \frac{400}{10} = 40$ أى يتم اختيار عشر عينات منتظمة كل منها (1) من (٤٠) .
أى أن لدينا (a) من العينات التى نريد اختيارها حيث حجم كل منها ($n' = \frac{n}{a}$) وطول فترة كل منها ($K' = \frac{N}{n'}$)

ويكون مقدر الوسط الحسابى للمجتمع (μ) مساوياً لـ :

$$\hat{\mu} = \frac{1}{a} \sum_{i=1}^{a} \bar{x}_i \qquad .... (7 - 19)$$

حيث : a عدد العينات المتكررة .

إن $\bar{x}_i$ هو الوسط الحسابى للعينة المنتظمة رقم (i) :

$$\bar{x}_i = \frac{1}{n'} \sum_{j=1}^{n'} \bar{x}_{ij} \qquad .... (7 - 20)$$

ويصبح مقدر تباين هذا المتوسط :

$$\hat{V}(\hat{\mu}) = \frac{N - n}{N} \times \frac{\sum_{i=1}^{a} (\bar{x}_i - \hat{\mu})^2}{a\,(a - 1)} \qquad .... (7 - 21)$$

أما مقدر القيمة الكلية فيصبح :

$$\hat{T} = N\,\hat{\mu} = \frac{N}{a}\sum_{i=1}^{a} \bar{x}_i \qquad \text{.... (7 - 22)}$$

ومقدر تباين القيمة الكلية يساوى :

$$\hat{V}(\hat{T}) = N^2\, V(\hat{\mu}) \qquad \text{.... (7 - 23)}$$

أى

$$\hat{V}(\hat{T}) = N^2\,\frac{N-n}{N}\,\frac{\sum_{i=1}^{a}(\bar{x}_i - \hat{\mu})^2}{a\,(a-1)} \qquad \text{.... (7 - 24)}$$

ويمكننا إهمال معامل تصحيح المجتمع $\frac{N-n}{N}$ عندما يكون حجم المجتمع (N) كبيراً .

تطبيق (٧ – ٩) :

يرغب أحد المصانع فى اختيار عينة منتظمة واحد من خمسة من علب إحدى السلع التى ينتجها البالغ عددها (٦٠) علبة . وقد تقرر اختيار عينة منتظمة متكررة من (٤) عينات لتقدير متوسط وزن العلبة وإجمالى وزن الإنتاج بمستوى معنوية (٠,٠٥) . المطلوب :

١ – توضيح كيفية اختيار وحدات العينة المنتظمة المتكررة .

٢ – تقدير متوسط وزن العلبة وإجمالى وزن العلب إذا كانت أوزان العلب كما يلى (بالكيلو غرام) :

رقم العلبة	الوزن	رقم العلبة	الوزن	رقم العلبة	الوزن	رقم العلبة	الوزن
1	10	16	12	31	13	46	14
2	11	17	13	32	14	47	16
3	12	18	14	33	18	48	19
4	10	19	11	34	20	49	19
5	13	20	14	35	13	50	14
6	14	21	15	36	13	51	14
7	16	22	17	37	16	52	17
8	16	23	17	38	17	53	18
9	18	24	19	39	19	54	20
10	16	25	17	40	16	55	15
11	17	26	18	41	14	56	13
12	16	27	18	42	12	57	13
13	19	28	20	43	16	58	16
14	10	29	11	44	14	59	13
15	11	30	12	45	19	60	18

الحـــل :

١ – إذا أردنا اختيار عينة منتظمة واحدة يكون لدينا

$$N = 60 \ , K = 5, n = \frac{60}{5} = 12$$

أى نختار (١٢) علبة حيث نختار من العلب الخمس الأولى رقمًا عشوائيًا وليكن (٣) ثم نضيف إليه طول الفترة بالتالى فتكون أرقام العلب المختارة :

3 8 13 18 23 28 33 38 43 48 53 58

وبالتالى نستخدم الصيغ الموضحة فيما سبق عند تقدير متوسط المجتمع والقيمة الكلية لعينة منتظمة واحدة .

٢ – يمكننا تقدير متوسط المجتمع باختيار عينة منتظمة متكررة مكونة من (٤) عينات بحيث يكون مجموع أحجامها يساوى (١٢) وحدة حيث أحجامها وطول فترتها كما يلى :

عدد العينات المتكررة $a = 4$

حجم العينة المتكررة $n' = \frac{12}{4} = 3$

طول الفترة $K^1 = \frac{N}{n'} = \frac{60}{3} = 20$

أى نختار (٤) عينات منتظمة متكررة حجم كل منها (٣) علب وكل منها واحد من (٢٠) . لذا نختار من الفترة الأولى التى أرقامها من ١ إلى ٢٠ أربعة أرقام تمثل هذه الأرقام العشوائية رقم المفردة الأولى لكل عينة . فإذا كانت هذه الأرقام هى على التوالى 6 , 14 , 17 ,12 تكون أرقام وحدات (العلب) العينات المختارة وأوزانها ومتوسطاتها :

	أرقام الوحدات	أوزان العلب	المتوسط
العينة الأولى	12, 32, 52	16, 14, 17	15.667
العينة الثانية	17, 37, 57	26, 16, 13	18.333
العينة الثالثة	14, 34, 54	10, 20, 20	16.667
العينة الرابعة	6, 26, 46	14, 18, 14	15.333

ويكون تقدير متوسط المجتمع

$$\hat{\mu} = \sum_{i=1}^{a} \frac{\bar{x}_i}{a}$$

$$= (15.667 + 18.333 + 16.667 + 15.333) / 4$$

$$= 66 / 4 = 16.5$$

– ويكون تقدير تباين تقدير متوسط المجتمع

$$\hat{V}(\hat{\mu}) = \frac{N - n}{N} \; \frac{\sum_{i=1}^{a} (\bar{x}_i - \hat{\mu})^2}{a(a-1)}$$

$$= \frac{60-12}{60} \times \frac{(15.667-16.5)^2+(18.333-16.5)^2+(16.667-16.5)^2+(15.333-16.5)^2}{4(4-1)}$$

$$= \frac{48}{60} \times \frac{5.4432}{12} = \frac{261.274}{720}$$

$$= 0.36288$$

– ويكون تقدير متوسط المجتمع بمستوى ثقة (٩٥٪) يساوى :

$$\hat{\mu} \pm t_{(1-\alpha/2, n-1)} \sqrt{\hat{V}(\hat{\mu})}$$

$$16.5 \pm 2.201 \times \sqrt{0.36288}$$

$$= 16.5 \pm 1.326$$

ويكون حدا الثقة باحتمال ٩٥٪ هما :

$$15.174 \leq \mu \leq 17.826$$

أى يتراوح متوسط وزن العلبة بين ١٥,١٧٤ كلغ و١٧,٨٢٦ كلغ وذلك بمستوى ثقة ٩٥٪ .

ويتراوح إجمالى وزن العلب بمستوى ثقة ٩٥٪ بين

$$15.174 \times 60 \leq T \leq 17.826 \times 60$$

$$910.44 \leq T \leq 1069.56$$

أى يتراوح بين ٩١٠,٤٤٠ كلغ و ١٠٦٩,٥٦٠ كلغ بمستوى ثقة ٩٥٪ .

تطبيق (٧ – ١٠) :

مجتمع من الموظفين مكون من (١٤) موظفًا كانت رواتبهم الشهرية (بآلاف الريالات) كما يلى :

٢ ، ٣ ، ٣ ، ٥ ، ٤ ، ٣ ، ٢ ، ٥ ، ٤ ، ٣ ، ٤ ، ٣ ، ٢ ، ٤

نريد اختيار عينة حجمها (٥) موظفين بالأسلوب المنتظم

المطلوب :

١ – توضيح كيفية اختيار العينة المنتظمة وما هى العينات الممكن سحبها ؟

٢ – إثبات أن تقدير متوسط العينة المنتظمة هو تقدير غير متحيز لمتوسط المجتمع .

٣ – استخراج تباين تقدير متوسط المجتمع وتقديره باستخدام بيانات العينة الأولى الممكن سحبها .

٤ – تقدير متوسط الراتب للموظفين وإجمالي رواتبهم بمستوى ثقة (٩٥٪) .

الحــل :

١ – نحسب طول الفترة (K) حيث

$$K= \frac{N}{n} = \frac{14}{5} = 3$$

وتكون العينة المختارة إحدى العينات الممكن سحبها التي مفرداتها :

العينة الأولى 2, 5, 2, 3, 2

العينة الثانية 3, 4, 5, 4, 4

العينة الثالثة 3, 3, 4 , 3

ويلاحظ أن حجم العينة الثالثة الممكن سحبها هو (٤) مفردات لأن (N) ليس من مضاعفات حجم العينة ، لذا يمكن في هذه الحالة إضافة الوحدة الأولى لتصبح العينة الثالثة الممكن سحبها 3, 3, 4, 3, 2 .

٢ – لإثبات أن متوسط العينة المنتظمة هو مقدر غير متحيز لمتوسط المجتمع ، نعلم أن

$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$= (2 + 3 + ... + 4) / 14$$

$$= 3.357$$

ونريد إثبات أن

$$E(\overline{x}_{sy}) = \frac{1}{K} \sum_{i=1}^{K} (\overline{x}_i) = \mu$$

$$(i = 1 , 2 , 3 \ (K = 3))$$

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij}$$

$$\bar{x}_{sy1} = \frac{14}{5} = 2.8 \text{ , } \bar{x}_{sy2} = \frac{20}{5} = 4 \text{ , } \bar{x}_{sy3} = \frac{13}{4} = 3.25$$

$$E(\bar{x}_{sy}) = \frac{1}{3}(2.8 + 4 + 3.25)$$

$$= \frac{10.05}{3} = 3.35$$

$$= \mu$$

أى أن الوسط الحسابى لعينة منتظمة هو مقدر غير متحيز لمتوسط المجتمع . (الفرق البسيط يعود بسبب اختلاف حجم العينة الأخيرة عن العينات الأخرى) .

٣ - يوجد عدة طرق لاستخراج تباين تقدير متوسط المجتمع وتقديره :

أ - نظراً لمعرفة العينات الممكن سحبها ، يكون

$$V(\bar{x}_{sy}) = \frac{1}{K} \sum_{i=1}^{K} (\bar{x}_i - \mu)^2$$

$$= \frac{1}{3}\left[(2.8 - 3.357)^2 + (4 - 3.357)^2 + (3.25 - 3.357)^2\right]$$

$$= 0.24$$

ب - نستخدم الصيغة التالية :

$$V(\bar{x}_{sy}) = \frac{N-n}{N} S^2 - \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

حيث

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (x_{ij} - \mu)^2$$

$$= \frac{1}{13}\left[(2 - 3.357)^2 + \dots + (3 - 3.357)^2 + \dots + (3 - 3.357)^2\right]$$

$$= 1.01$$

ويكون الحد الأول من الطرف الأيمن :

$$\frac{N-1}{N} S^2 = \frac{13}{14} \times 1.01 = 0.94$$

أما الحد الثاني من الطرف الأيمن فيساوى :

$$\frac{1}{14}\Big[(2-2.8)^2 + (5-2.8)^2 + \ldots\ldots\ldots + (3-4)^2 + (4-4)^2 + \ldots\ldots\ldots$$

$$+ (3-3.25)^2 + \ldots\ldots + (3-3.25)^2 \Big]$$

ويكون

$$V(\overline{x}_{sy}) = 0.94 - 0.68 = 0.26$$

والخطأ المعيارى :

$$\sigma_{\overline{x}_{sy}} = \sqrt{V(\overline{x}_{sy})}$$

$$= \sqrt{0.26} = 0.51$$

ج – باستخدام بيانات العينة الأولى الممكن سحبها ، يكون مقدر تباين تقدير متوسط المجتمع :

$$\hat{V}(\overline{x}_{sy}) = \frac{s^2}{n}(1-f)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s^2 = \frac{1}{5-1}\Big[(2-2.8)^2 + \ldots\ldots + (2-2.8)^2\Big]$$

$$= \frac{6.8}{4} = 1.7$$

ويكون

$$\hat{V}(\overline{x}_{sy}) = \frac{1.7}{5}\left(1 - \frac{5}{14}\right) = 0.22$$

$$\hat{\sigma}_{\overline{x}_{sy}} = \sqrt{0.22} = 0.47$$

- حدا الثقة للوسط الحسابى

$$\bar{x}_{sy} \mp t_{(1-\alpha/2, n-1)} \hat{\sigma}_{\bar{x}_{sy}}$$

$$2.8 \mp 2.776 \times 0.47$$

$$= 2.8 \mp 1.30$$

ويكون الحد الأدنى ١,٥ والحد الأعلى ٤,١ أى أن متوسط المجتمع يتراوح بين ١,٥ و٤,١ وذلك بمستوى ثقة ٩٥ ٪ أى :

$$1.5 \le \mu \le 4.1$$

أما تقدير القيمة الكلية للإنفاق الشهرى :

$$\hat{T} = N\ \bar{x}_{sy} = 14 \times 2.8 = 39.2$$

$$\hat{V}\ (\hat{T}) = N^2\ V\,(\bar{x}_{sy})$$

$$= 14^2 \times 0.22 = 43.12$$

$$\hat{\sigma}_{\hat{T}} = \sqrt{43.12} = 6.65$$

ويكون حدا الثقة للقيمة الكلية :

$$\hat{T} \mp t_{(1-\alpha/2, n1)}\ \hat{\sigma}_{\hat{T}}$$

$$= 39.2 \mp 2.776 \times 6.56$$

$$= 39.2 \mp 18.2$$

ويكون الحد الأدنى (٢١) ويكون الحد الأعلى (٥٧,٤) بمستوى ثقة ٩٥٪ أى أن :

$$21 \le T \le 57.4$$

الفصل
الثامن

المعاينة العنقودية البسيطة

Simple Cluster Sampling

٨ - ١ تعريف المعاينة العنقودية البسيطة :

عندما يكون حجم المجتمع المراد دراسته كبيراً ، فإن استخدام أحد أنواع العينات السابقة يتطلب إعداد أو توافر إطار جميع الوحدات ومن ثم اختيار وحدات العينة المناسبة ويتطلب ذلك إمكانات بشرية ومالية كبيرة .

لذا يفضل بعض الباحثين دراسة جزء من المجتمع بدقة عالية للحصول على تقديرات جيدة تمثل معالم المجتمع الإحصائى أى نحتاج فقط لإطار الجزء الذى يتم دراسته فقط ، ونكون بذلك قد سحبنا عينة دون الحاجة لإطار جميع الوحدات ووفرنا الوقت والمال والجهد .

تتلخص طريقة اختيار وحدات المعاينة العنقودية البسيطة فى تقسيم المجتمع الإحصائى إلى وحدات أولية (تسمى العناقيد الأولية أو الابتدائية (Primary Clusters)) وكل عنقود منها مؤلف من عدد الوحدات الإحصائية التى تسمى "الوحدات المشاهدة" . يتم اختيار عدد من العناقيد الأولية باستخدام إحدى طرق الاختيار العشوائى ويتم حصر العناقيد المختارة حصراً شاملاً .

ويمكننا تعريف المعاينة العنقودية البسيطة بأنها "معاينة عشوائية بسيطة تكون فيها كل وحدة معاينة مجموعة (أوعنقود) من الوحدات المشاهدة" .

وتعد تكلفة المعاينة العنقودية أقل من تكلفة المعاينة العشوائية البسيطة أو المعاينة الطبقية أو المنتظمة . إن استخدام هذا النوع من المعاينات يؤدى إلى توفير التكاليف بسبب عدم وجود مسافات كبيرة بين وحدات العينة لأنها تقع بجانب بعضها ضمن العنقود الواحد الذى يتم حصر جميع وحداته حصراً شاملاً .

وهكذا يمكننا القول إنه يمكن استخدام المعاينة العنقودية البسيطة بشكل مناسب للحصول على البيانات بأقل تكلفة فى الحالتين التاليتين :

- عندما يكون إطار الوحدات الإحصائية الذى يتضمن أسماءها وعناوينها غير متوافر أو أن إعداده يتطلب نفقات ضخمة .
- ضخامة نفقات الحصول على البيانات من الوحدات نتيجة انتشار الوحدات ووجود مسافات كبيرة بينها .

تسمى أحياناً المعاينة العنقودية البسيطة ، المعاينة العنقودية ذات المرحلة الواحدة (Single Stage Cluster Sampling) وذلك للتمييز بينها وبين المعاينة العنقودية ذات المرحلتين والمعاينة العنقودية ذات المراحل المتعددة التى سنتعرض لهما فى الفصل القادم .

٨ – ٢ طريقة اختيار العينة العنقودية البسيطة :

إن الخطوة الأولى الواجب اتباعها لاختيار وحدات العينة العنقودية البسيطة هى تحديد العناقيد الأولية (المجموعات الابتدائية) التى سنقوم باختيار عدد منها بشكل عشوائى . إن الوحدات التى يتضمنها كل عنقود أو مجموعة غالبًا ما يكون لها خصائص متشابهة ومتقاربة مع بعضها تشابهًا وتقاربًا طبيعيًا . وبعبارة أخرى يمكن القول إن قياس أية وحدة فى العنقود قد يكون مرتبطًا بشكل قوى مع قياس الوحدة الأخرى ، لذا فإن المعلومات المتعلقة بمعالم المجتمع ، قد لا تزداد بشكل ملحوظ إذا أخذت بيانات أخرى جديدة ضمن العنقود الواحد ، وجمع البيانات من عدد كبير من الوحدات ضمن العنقود سيؤدى إلى زيادة التكاليف . ومع ذلك فإن الاهتمام يجب أن يركز على الحالات التى تكون فيها الوحدات ضمن العنقود الواحد تختلف من وحدة لأخرى . فى مثل هذه الحالات فإن اختيار عدد قليل من العناقيد الكبيرة سيعطى تقدير «جيد» لمعلمة المجتمع كالوسط الحسابى . ولكن أفضل العناقيد هى التى تعطى تقديراً للخاصية التى ندرسها بأصغر انحراف معيارى ، أى أنه كلما صغر حجم العنقود كلما زادت دقة التقدير لعينة ذات حجم محدد ، وذلك لأنه سيتم حصر العناقيد المختارة حصراً شاملاً ، وازدياد عدد وحدات العنقود سيؤدى إلى زيادة الانحراف المعيارى .

وبعد تحديد عدد العناقيد الابتدائية ، يتم اختيار عينة عشوائية بسيطة من هذه العناقيد باستخدام إحدى طرق السحب العشوائى ، ويتم حصر كل من العناقيد المختارة حصراً شاملاً وتكون مفردات العينة العنقودية البسيطة هى القيم الإجمالية للعناقيد المختارة ، أى كل مفردة هى عبارة عن القيمة الإجمالية للعنقود الذى تم اختياره ، وتكون لدينا مفردات عددها يساوى عدد العناقيد المختارة .

ويجب علينا عند دراسة المعاينات العنقودية الانتباه إلى عدد الوحدات التى يتكون منها كل عنقود (حجم العنقود) إذ هناك العناقيد ذات الحجم المتساوى (Equal size clusters) والعناقيد ذات الحجم غير المتساوى (Unequal size clusters) .

تطبيق (٨ – ١) :

لتوضيح طريقة اختيار العينة العنقودية البسيطة ، لنفترض أننا نرغب فى تقدير عدد السكان ومتوسط حجم الأسرة فى إحدى المدن ولا يتوافر لدينا إطار الأسر أى قائمة بأسماء رؤساء الأسر وعناوينهم . كما أن تكاليف إعداد الإطار ضخمة ، خاصة أن عدد الأسر كبير ويتطلب أيضًا وقتًا كبيراً وإمكانات بشرية كبيرة . نستخدم فى هذه الحالة العينة العنقودية البسيطة ، إذ يمكن تقسيم المدينة إلى مجموعات (عناقيد) حسب معايير معينة ، مثلاً : نستخدم التقسيم الشائع الاستخدام للمدينة أى الأحياء كمعيار ، وبذلك يكون المجتمع لدينا

مكونًا من عناقيد ابتدائية (أو أولية) عددها يساوى عدد الأحياء ، ونقوم باختيار عدد من العناقيد (الأحياء) باستخدام طرق السحب العشوائى ، وتكون العينة العنقودية مكونة من العناقيد المختارة أى من الأحياء المختارة . ونقوم بإعداد إطار فقط للأحياء المختارة وحصرها حصراً شاملاً ، ويكون عدد المفردات فى هذه الحالة يساوى عدد العناقيد المختارة وقيمة كل منها يساوى عدد سكان الحى .

ويتم تحديد عدد العناقيد المختارة (حجم العينة) باستخدام الصيغة المناسبة التى سنتطرق إليها فى الصفحات القادمة .

٨ - ٢ رموز ومصطلحات :

ليكن لدينا مجتمع إحصائى ونرغب فى اختيار عينة عنقودية بسيطة ، فإننا نستخدم الرموز التالية :

M عدد عناقيد المجتمع .

m عدد العناقيد المختارة المكونة للعينة .

N_i عدد الوحدات (المفردات) فى العنقود (i) فى المجتمع (حيث $(i = 1,2, ---, M)$) .

N عدد الوحدات (المفردات) التى تحتويها جميع عناقيد المجتمع .

n_i عدد الوحدات (المفردات) التى يحتويها العنقود (i) فى العينة .

n عدد الوحدات (المفردات) التى تحتويها جميع عناقيد العينة .

x_i قيمة العنقود أى إجمالى قيمة مفردات العنقود (i) .

x_{ij} قيمة المفردة (المشاهدة) (j) فى العنقود (i) .

وبالتالى يمكننا تمثيل عناقيد المجتمع كما يلى :

العنقود	1	2	i	----	M
	x_{11}	x_{21}	x_{i1}	----	x_{M1}
	x_{12}	x_{22}	x_{i2}	----	x_{M2}
	x_{13}	x_{23}	x_{i3}	----	x_{M3}
	.	.	.	----	.
	x_{1j}	x_{2j}	x_{ij}	----	x_{Mj}
	.	.	.	----	.
	x_{1N_1}	x_{2N_2}	x_{iN_i}	----	x_{MN_M}
قيمة العنقود	χ_1	χ_2	χ_3	----	χ_M
متوسط العنقود فى المجتمع	μ_1	μ_2	μ_3		μ_M
عدد مفردات المجتمع	N_1	N_2	N_3	----	N_M

ويمكننا القول إن :

– عدد مفردات جميع عناقيد المجتمع (N) يساوى :

$$N = \sum_{i=1}^{M} N_i \qquad \text{.... (8 - 1)}$$

وهنا نميز بين حالتين :

أ – عدد مفردات كل عنقود متساوٍ ، أى أن

$N_1 = N_2 = \text{---} = N_i = \text{---} = N_M$

وبالتالى يكون عدد مفردات المجتمع يساوى

$N = M\, N_i$

حيث N_i تمثل حجم العنقود (i) ويساوى حجم أى عنقود لأنها متساوية من حيث الحجم .

ب – عدد مفردات كل عنقود يختلف من عنقود لآخر ، أى أن عدد مفردات المجتمع يساوى (N) كما هو موضح فى الصيغة (1 - 8) وبالتالى يمكننا القول إن متوسط حجم العنقود فى المجتمع يساوى :

$$\overline{N} = \frac{1}{M} \sum_{i=1}^{M} N_i \qquad \text{.... (8 - 2)}$$

أى أن

$$\overline{N} = \frac{N}{M}$$

– يمكننا استخراج عدد مفردات العينة بالطريقة السابقة نفسها إذ نجد أن عدد عناقيد العينة يساوى (m) عنقوداً فإذا كان حجم كل منها متساوٍ ، أى أن :

$n_1 = n_2 = \text{-----} = n_i = \text{-----} = n_m$

وبالتالى فإن :

$$n = \sum_{i=1}^{m} n_i$$

$n = m\, n_i$

حيث (n_i) يساوى حجم العنقود (i) وهو متساوٍ لجميع عناقيد العينة .

أما إذا كانت أحجام عناقيد العينة غير متساوية فإن عدد مفردات العينة التى تتضمن m عنقوداً يساوى :

$$n = \sum_{i=1}^{m} n_i$$

وبالتالى يكون متوسط حجم عنقود العينة :

$$\bar{n} = \frac{1}{m}\sum_{i=1}^{m} n_i \qquad \text{.... (8 - 3)}$$

أى أن :

$$\bar{n} = \frac{n}{m}$$

٨ – ٤ تقدير أهم معالم المجتمع :

٨ – ٤ – ١ تقدير الوسط الحسابى للمجتمع والقيمة الكلية للمجتمع :

إن العينة العنقودية البسيطة هى عبارة عن عينة عشوائية بسيطة تتضمن (m) وحدة (عنقود) قيمة المفردة (i) فيها هى (x_i) (حيث $i = 1, 2, ---, m$) أى إجمالى قيمة العنقود ، أى أن :

$$x_i = \sum_{j=1}^{n_i} x_{ij}$$

وبالتالى يكون متوسط قيمة المفردة فى العنقود (i) فى العينة :

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (8 - 4)}$$

ويساوى :

$$\bar{x}_i = \frac{x_i}{n_i} = \frac{x_i}{N_i} = \mu_i$$

لأن مفردات العنقود (i) فى المجتمع هى مفردات العنقود (i) نفسها فى العينة العنقودية البسيطة أى أن $X_i = x_i$ لذا سنستخدم x_i عند دراستنا لهذه العينة . وحيث لدينا (m) قيمة كل منها يمثل القيم الإجمالية للعينة .

$$x_1 , x_2 , x_3 , \ldots , x_m$$

$$x = \sum_{i=1}^{m} x_i$$

لذا يكون تقدير متوسط المجتمع $(\hat{\mu})$ من بيانات عينة عنقودية ولنرمز له بالرمز $(\bar{x})$ أى تقدير متوسط قيمة المفردة :

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{m} x_i$$

أى أن :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij} \qquad \ldots (8 - 5)$$

أى يساوى : $\bar{x} = \frac{x}{n}$

– أما متوسط قيمة العنقود فى العينة فيساوى :

$$\bar{x}_{cl} = \frac{1}{m} \sum_{i=1}^{m} x_i \qquad \ldots (8 - 6)$$

حيث لدينا (m) عنقوداً . وسنستفيد من هذه الصيغ والرموز فى الفصل القادم عند دراسة المعاينة العنقودية ذات المرحلتين أو ذات المراحل المتعددة .

– إن مقدر القيمة الكلية للمجتمع من بيانات عينة عنقودية يساوى :

$$\hat{T} = N \bar{x}$$

$$\hat{T} = \frac{N}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij} \qquad \ldots (8 - 7)$$

وعندما يكون حجم المجتمع (N) مجهولاً نستخدم الصيغة التالية :

$$\hat{T} = \frac{M}{m} \sum_{i=1}^{m} x_i \qquad .... (8 - 8)$$

أى أن :

$$\hat{T} = M \; \bar{x}_{c1}$$

ويعد مقدر متوسط المجتمع من بيانات عينة عنقودية الصيغة (6 - 8) مقدراً غير متحيز لمتوسط المجتمع ، كذلك يعد مقدر القيمة الكلية للمجتمع من بيانات هذه العينة باستخدام الصيغة (7 - 8) أو (8 - 8) مقدراً غير متحيز للقيمة الكلية للمجتمع .

تطبيق (٨ – ٢) :

تتكون إحدى الوزارات من (٢٠) إدارة رئيسية يبلغ عدد موظفيها (١٤٠) موظفًا . وقد تم اختيار عينة عنقودية من (٧) إدارات وذلك لتقدير متوسط الإنفاق الشهرى (بالآلاف) للموظف ، وكانت البيانات المستخرجة كما يلى :

رقم الإدارة (العنقود)	عدد الموظفين	إجمالى الرواتب
١	٥	٢٠
٢	٧	٢١
٣	٨	٣٠
٤	٥	٢٥
٥	٦	٢٧
٦	٦	٢٨
٧	٥	٢٤
المجموع	٤٢	١٧٥

المطلوب :

١ – تقدير متوسط الإنفاق الشهرى للموظف .

٢ – تقدير إجمالى الإنفاق الشهرى للموظفين فى هذه الوزارة .

الحـل :

من البيانات نجد أن :

– عدد عناقيد المجتمع (عدد إدارات المجتمع) $M = 20$.

– عدد العناقيد المختارة فى العينة (عدد الإدارات المختارة) $m = 7$.

– عدد مفردات المجتمع $N = 140$.

– عدد مفردات العينة $n = \sum_{i=1}^{7} n_i = 42$

– إجمالى رواتب كل عنقود (إدارة) يساوى :

$$x_1 = 20,\ x_2 = 21, \text{------},\ x_m = 24$$

وبالتالى فإن إجمالى الإنفاق من بيانات العينة يساوى :

$$x = \sum_{i=1}^{m} x_i$$

$$= 20 + 21 + \text{----} + 24 = 175$$

– إن تقدير متوسط العنقود :

$$\bar{x}_{cl} = \frac{x}{m} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

$$= \frac{1}{7}\, 175 = 25$$

أى أن متوسط إجمالى إنفاق موظفى كل إدارة من بيانات العينة يساوى (٢٥) ألف ريال ، وهو تقدير غير متحيز لمتوسط إجمالى إنفاق موظفى الإدارة الواحدة ، أما تقدير متوسط الإنفاق الشهرى للموظف فى هذه الوزارة فيساوى :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{1}{n} \sum_{i=1}^{m} x_i$$

$$= \frac{1}{42} \text{ x } 175 = 4.16667$$

أى (٤١٦٦,٦٧) ريالاً وهو الطلب الأول .

– أما تقدير إجمالى إنفاق موظفى الوزارة فيساوى :

$$\hat{T} = \frac{N}{n} \sum_{j=1}^{m} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{140}{42} \text{ x } 175 = 583.330$$

أى أن تقدير إجمالى الإنفاق الشهرى لموظفى الوزارة يساوى (٥٨٣٣٣٠) ريالاً وهو تقدير غير متحيز لإجمالى إنفاق موظفى الوزارة (إجمالى إنفاق المجتمع) .

– ويمكن تقدير إجمالى إنفاق الوزارة بطريقة أخرى خاصة عندما يكون حجم المجتمع (N) مجهولاً وذلك باستخدام متوسط العنقود ($\bar{x}_{c1}$) حيث نجد أن تقدير القيمة الإجمالية للمجتمع يساوى :

$$\hat{T} = M \quad \bar{x}_{c1}$$

وحسب بيانات التطبيق نجد أن :

$$\hat{T} = 20 \text{ x } 25$$

$$= 500$$

أى أن تقدير إجمالى إنفاق موظفى الوزارة هو (٥٠٠٠٠٠) ريال وهى قيمة قريبة من القيمة التى حصلنا عليها فيما سبق وتستخدم هذه الطريقة عندما يكون حجم المجتمع مجهولاً .

٨ - ٤ - ٢ تقدير تباين تقدير متوسط المجتمع وتقدير تباين تقدير القيمة الكلية للمجتمع :

إن الصيغة المستخدمة لتقدير تباين تقدير متوسط المجتمع $\hat{V}(\bar{x})$ تساوى :

$$\hat{V}(\bar{x}) = \frac{M - m}{M m \bar{N}^2} \frac{\sum_{i=1}^{m} (x_i - \bar{x} n_i)^2}{m - 1} \quad .... (8 - 9)$$

حيث $(\bar{N} = \frac{N}{M})$

إن $\hat{V}(\bar{x})$ الموضحة فى الصيغة (9 - 8) هى مقدر متحيز ، لكنه مقدر جيد لـ $V(\bar{x})$ إذا كان عدد العناقيد المختارة (m) كبيراً (مثلاً $m \geq 20$) . إن هذا التحيز يتلاشى إذا كانت أحجام العناقيد $(n_1, n_2, ----, n_M)$ متساوية .

أما مقدر تباين تقدير القيمة الكلية للمجتمع فيساوى إحدى الصيغتين التاليتين :

– نستخدم الصيغة التالية إذا قدرنا متوسط المجتمع باستخدام الصيغة (5 - 8) وذلك عندما يكون حجم المجتمع (N) معلوماً :

$$\hat{V}(\hat{T}) = \hat{V}(N \bar{x})$$

$$= N^2 \hat{V}(\bar{x})$$

ويساوى :

$$\hat{V}(\hat{T}) = M^2 \frac{(M - m)}{M m} \frac{\sum_{i=1}^{m} (x_i - \bar{x} n_i)^2}{m - 1} \quad .... (8 - 10)$$

أما عندما يكون حجم المجتمع (N) غير معلوم فإننا نستخدم الصيغة التالية والتى تستخدم إذا قدرنا متوسط المجتمع باستخدام الصيغة (6 - 8) :

$$\hat{V}(\hat{T}) = \hat{V}(M \bar{x}_{c1}) = M^2 \hat{V}(\bar{x}_{c1})$$

أى يساوى :

$$\widehat{V}\ (\widehat{T}\ ) = M^2\ \frac{(M - m\ )}{M\ m}\ \frac{\sum_{i=1}^{m} (x_i - \bar{x}_{cl})^2}{m - 1} \quad .... (8 - 11)$$

٨ – ٥ حدود الثقة لتقدير متوسط المجتمع وتقدير القيمة الكلية للمجتمع :

إن حدى الثقة لتقدير متوسط المجتمع بمستوى ثقة $(1-\alpha)\%$:

$$\bar{x} \mp Z_{(1-\alpha/2)} \sqrt{\widehat{V}\ (\bar{x})} \quad .... (8 - 12)$$

حيث $\widehat{V}\ (x)$ هو تقدير تباين متوسط المجتمع المحسوبة باستخدام إحدى الصيغ السابقة و(Z) هى القيمة المقابلة فى جدول التوزيع الطبيعى باحتمال $(1 - \alpha/2)$ وعندما يكون حجم العينة صغيراً نستخدم القيمة المقابلة فى جدول توزيع (t) .

أما حدا الثقة لتقدير القيمة الكلية للمجتمع فهما :

$$\widehat{T} \mp Z_{(1-\alpha/2)} \sqrt{\widehat{V}\ (\widehat{T})} \quad .... (8 - 13)$$

حيث $\widehat{V}\ (\widehat{T})$ هو مقدر تباين تقدير القيمة الكلية المحسوبة باستخدام إحدى الصيغ السابقة .

ولتوضيح كيفية حساب حدود الثقة نورد التطبيق التالى :

تطبيق (٨ – ٣) :

باستخدام بيانات التطبيق (٨ – ٢) ما هى حدود الثقة لتقدير متوسط المجتمع وتقدير القيمة الكلية بمستوى ثقة (٩٥٪) ؟

الحل :

– نستخدم الصيغة (11 - 8) لاستخراج تقدير تباين تقدير المتوسط التى تتطلب حساب المقدار :

$$\sum_{i=1}^{7} (x_i - \bar{x} n_i)^2 = \sum x_i^2 - 2\bar{x}\sum_{i=1}^{m} x_i n_i + \bar{x}^2 \sum_{i=1}^{m} n_i^2$$

لذا ننظم الجدول التالى :

رقم العنقود	1	2	3	4	5	6	7	المجموع
قيمة العنقود (x_i)	20	21	30	25	27	28	24	175
عدد مفردات العنقود (n_i)	5	7	8	5	6	6	5	42
متوسط العنقود $(\bar{x}_i)$	4	3	3.75	5	4.5	4.7	4.80	---
n^2	25	49	64	25	36	36	25	260
$x_i n_i$	100	147	240	125	162	168	120	1062
x_i^2	400	441	900	625	729	784	576	4455

$$\sum_{i=1}^{7} (x_i - \bar{x} n_i)^2 = \sum x^2 - 2\bar{x}\sum_{i=1}^{m} x_i n_i + \bar{x}^2 \sum_{i=1}^{m} n_i^2$$

$$= 4455 - 2 \times 4.16667 \times 1062 + (4.16667)^2 \times 260$$

$$= 4455 - 8850 + 4513.90$$

$$= 118.9$$

وبالتالي يكون تقدير تباين تقدير متوسط المجتمع :

$$\hat{V}(\bar{x}) = \frac{M - m}{M m \bar{N}^2} \frac{\sum (x_i - \bar{x} n_i)^2}{m - 1}$$

$$= \frac{20 - 7}{20 \times 7 \times \left(\frac{140}{20}\right)^2} \times \frac{118.9}{7 - 1}$$

$$= \frac{1545.7}{41160} = 0.0376$$

ويكون حد الثقة لتقدير متوسط المجتمع بمستوى ثقة ٩٥٪ :

$$4.16667 \pm 1.96 \sqrt{0.0376}$$

$$= 4.16667 \pm 0.38$$

أى أن :

$$3.78667 \leq \mu \leq 4.54667$$

أى سيتراوح متوسط إنفاق الموظف بين (٣٧٨٦,٦٧) ريالاً و(٤٥٤٦,٦٧) ريالاً وذلك بدرجة ثقة (٩٥٪) :

تقدير تباين تقدير القيمة الكلية للمجتمع يساوى :

$$\hat{V}(\hat{T}) = M^2 \frac{M - m}{M m} \frac{\sum_{i=1}^{m} (x_i - \bar{x} n_i)^2}{m - 1}$$

$$= 20^2 \frac{(20 - 7)}{20 \times 7} \times \frac{118.9}{7 - 1}$$

$$= \frac{618280}{840} = 736.047$$

ويساوى أيضاً :

$$\hat{V}(\hat{T}) = \hat{V}(N \bar{x})$$

$$= N^2 V(\bar{x})$$

$$= (140)^2 \times 0.0376 = 736.96$$

ويعود الفرق بين التقدير للتقريب .

ويكون حدا الثقة لتقدير القيمة الكلية للمجتمع بمستوى ثقة ٩٥٪ .

$$\hat{T} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{T})}$$

$$583.330 \pm 1.96\sqrt{736.647}$$

$$583.330 \mp 53.1752$$

أى أن

$$530.1584 \le T \le 636.5052$$

أى أن إجمالى الإنفاق للموظفين سيتراوح بدرجة ثقة (٩٥٪) بين (٥٣٠١٥٨,٤) ريالاً و(٦٣٦٦٠٥,٢) ريالات .

ويمكننا استخدام الصيغة (10 - 8) لحساب $\hat{V}(\hat{T})$ إذا كان حجم المجتمع (N) غير معلوم .

٨ - ٦ تقديرات نسبة المجتمع وتباين نسبة المجتمع :

(Estimation of population proportion and variance)

كثيرًا ما يرغب الباحث فى تقدير نسبة المجتمع للذين يتصفون بخاصية معينة باستخدام المعاينة العنقودية البسيطة ، مثلاً قد نرغب فى تقدير نسبة الموافقين على إجراءات جديدة ستطبق على موظفى الوزارات . نقوم فى هذه الحالة ، باختيار عدد من الوزارات (العناقيد الأولية) عشوائيًا ثم نقوم بحصر هذه الوزارات المختارة حصرًا شاملاً وحساب عدد الذين يوافقون على هذه الإجراءات . فإذا رمزنا إلى عدد الذين يتصفون بالخاصية المدروسة (عدد الموافقين على الإجراءات مثلاً) فى العنقود (i) من عناقيد العينة بالرمز (a_i) ، يكون لدينا $a_1 , a_2 , \cdots , a_m$ نجد أن مقدر نسبة المجتمع ولنرمز له بالرمز p (أو $\hat{P}$) يساوى :

$$P = \frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{m} n_i} \qquad \text{.... (8 - 14)}$$

حيث (n_i) عدد وحدات العنقود (i) فى العينة حيث (i = 1 , 2 , ... m) أما تقدير تباين تقدير نسبة المجتمع $\hat{V}(p)$ فيساوى :

$$\hat{V}(\hat{P}) = \frac{M - m}{M\, m\, \overline{N}^{2}} \times \frac{\sum_{i=1}^{m}(a_i - p\, n_i)^2}{m - 1} \qquad \text{.... (8 - 15)}$$

حيث :

$$\sum_{i=1}^{m} (a_i - p\ n_i)^2 = \sum_{i=1}^{m} a_i^2 - 2\ p \sum_{i=1}^{m} a_i\ n_i + p^2 \sum_{i=1}^{m} n_i^2$$

وعندما يكون $\bar{N}_i$ غير معلوم ، نستخدم $(\bar{n}_i)$ حيث $(\bar{N}_i = \bar{n}_i)$ وتعد صيغة التباين رقم (8 - 15) مقدراً جيداً فقط عندما يكون عدد العناقيد المختارة (حجم العينة m) كبيراً (وليكن $m \geq 20$) . وتعد (p) الموضحة في الصيغة (8 - 14) مقدراً غير متحيز لنسبة المجتمع (P) كما يعد التباين الموضح في الصيغة (8 - 15) مقدراً غير متحيز لتباين نسبة المجتمع V (P) لأى حجم من العينات إذا كان حجم العناقيد المختارة متساوياً أى $(n_1 = n_2 = \cdots = n_m)$.

أما حدا الثقة لتقدير نسبة المجتمع بمستوى ثقة $(1-\alpha)\%$ فهما :

$$p \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(p)} \qquad \text{.... (8 - 16)}$$

ويمكن استخدام (t) عوضاً عن (Z) إذا كان حجم العينة (m) صغيراً .

أما تقدير إجمالي الذين يتصفون بخاصية معينة $(\hat{T}_a)$ فيساوى :

$$\hat{T}_a = N\ p \qquad \text{.... (8 - 17)}$$

ويمكن استخراج حدود الثقة باستخراج تباين تقدير المجموع الذى يساوى $N^2\ V(p)$.

تطبيق (٨ – ٤) :

يرغب أحد الباحثين في دراسة مستوى الخدمات المقدمة للمرضى في أحد المستشفيات الذى يتكون من (٥٠) قسماً . وقد اختيرت عينة مكونة من (٢٢) قسماً تم حصر أراء المرضى فيها حصراً شاملاً وكانت البيانات كما يلي :

عدد المرضى في المستشفى (٤٠٠) مريض .

عدد المرضى الذين اختيروا كعينة (٢٢٠) مريضاً منهم (١٨٠) مريضاً يرون أن مستوى الخدمات المقدمة جيد .

– ما هو تقدير نسبة المرضى في المستشفى الذين يرون أن مستوى الخدمات جيد بمستوى ثقة (٩٥٪) ؟

– ما هو تقدير إجمالي المرضى الذين يرون أن مستوى الخدمات جيد بمستوى ثقة (٩٥٪) ؟

علمًا بأن $\Sigma a_i^2 = 1280 , \Sigma a_i n_i = 8190 , \Sigma n_i^2 = 18720$

الحــل :

لدينا :

$$\sum a_i^2 = 1280 , \sum a_i \ n_i = 8190 . \sum n_i^2 = 18720$$

$$M = 50 , m = 22 , N = 400$$

$$n = 220 , \sum_{i=1}^{m} a_i = 180$$

حيث (a_i) تمثل عدد الذين يرون أن مستوى الخدمات جيد في العنقود (i) من العينة .
ويكون تقدير نسبة الذين يرون أن مستوى الخدمات جيد :

$$p = \frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{m} n_i}$$

$$= \frac{180}{220} = 0.8182$$

أما تقدير تباين نسبة المجتمع فيساوي :

$$\hat{V}(p) = \frac{M - m}{M m \bar{N}^2} \frac{\sum (a_i - p \, n_i)^2}{m - 1}$$

$$\bar{N} = \bar{n} = \frac{220}{22} = 10$$

$$\sum_{i=1}^{m} (a_i - p \ n_i)^2 = \sum a_i^2 - 2 \ p \sum_{i=1}^{m} a_i \ n_i + p^2 \sum_{i=1}^{m} n_i^2$$

$$= 1280 - (2 \times 0.8182 \times 8190) + (0.8182)^2 \times 18720$$

$$= 1280 - 13402 + 12532$$

$$= 410$$

ويكون :

$$\widehat{V}(p) = \frac{50 - 22}{50 \times 22 \times 10^2} \times \frac{410}{22 - 1}$$

$$= \frac{11480}{2310000} = 0.00496$$

$$\sqrt{\widehat{V}(p)} = 0.0705$$

ويكون حدا الثقة كما يلى (بمستوى ثقة ٩٥٪) :

$$0.8182 \mp 1.96 \times 0.0705$$

$$= 0.8182 \mp 0.13818$$

أى أن الحد الأدنى (0.68) والحد الأعلى (0.9564) أى أن نسبة الذين يرون أن مستوى الخدمات فى المستشفى جيد يتراوح بين هاتين النسبتين بمستوى ثقة ٩٥٪ أى :

$$0.68 \leq p \leq 0.9564$$

– أما تقدير إجمالى عدد الذين يرون أن مستوى الخدمات جيد فى المستشفى ولنرمز له بـ $(\widehat{T}_a)$.

$$\hat{T}_a = N p$$

$$= 400 \times 0.8182 = 327$$

ويتراوح هذا العدد بمستوى ثقة (٩٥٪) بين (٢٧٢) و(٣٨٣) لأن :

$$0.68 \times 400 \leq T_A \leq 0.9564 \times 400$$

$$272 \leq T_A \leq 383$$

حيث T_A يمثل عدد الذين يرون أن الخدمات جيدة فى المجتمع أى مرضى المستشفى .

٨ - ٧ تحديد حجم العينة :

تتأثر بيانات العينة العنقودية البسيطة بعدد العناقيد وحجم كل عنقود فيها . وقد كنا فيما سبق نركز على عملية اختيار عدد من العناقيد (m عنقوداً) من عناقيد المجتمع التى عددها (M) عنقوداً ، وذكرنا أن تقدير تباين متوسط المجتمع الصيغة (9 - 8) يساوى :

$$\hat{V}(\bar{x}) = \frac{M - m}{M\, m\, \bar{N}^2}\ (s_{cl}^2) \qquad \text{.... (8 - 18)}$$

$$s_{cl}^2 = \frac{\sum_{i=1}^{m} (x_i - \bar{x}\, n_i)^2}{m - 1}$$

إن التباين الفعلى لتقدير متوسط المجتمع ، يقتضى استخدام (σ_{cl}^2) أو (S_{cl}^2) عوضاً عن (s_{cl}^2) أى نستخدم تباين العنقود من المجتمع وليس من العينة أى تقديره (s_{cl}^2) ، ولكننا لا نعلم متوسط حجم العنقود $(\bar{N})$ وأيضاً لا نعلم (σ_{cl}^2) ، لذا نجد صعوبة فى تحديد حجم العينة اللازم للحصول على المعلومات اللازمة لمعلمة المجتمع . لذا نستخدم تقدير (σ_{cl}^2) و $(\bar{N})$ التى يمكن الحصول عليها من عينة استطلاعية أو اختيار عينة بشكل أولى وتقدير حجم العينة أى عدد العناقيد (m) .

لتحديد حجم العينة نستخدم حد خطأ التقدير (β) الذى يمكن اختياره من قبل الخبراء ، وهو يمثل الخطأ الأعظم الذى يقبلونه ويساوى :

$$\beta = Z_{(1-\alpha/2)} \sqrt{V(\bar{x})}$$

ونستخدم الصيغة التالية لتحديد حجم العينة المطلوب لتقدير متوسط المجتمع (μ) بخطأ تقدير (β) :

$$m = \frac{M\, \sigma_{cl}^2}{M\, D + \sigma_{cl}^2} \qquad \text{.... (8 - 19)}$$

وعندما يكون (σ_{cl}^2) مجهولاً نستخدم تقديره (s_{cl}^2) و $\left(D = \frac{\beta^2\ \bar{N}^2}{Z^2}\right)$ وفى حال عدم معرفة حجم المجتمع ومتوسط حجم العنقود ، نستخدم متوسط حجم العينة كتقدير له .

أما الصيغة الممكن استخدامها لتحديد حجم العينة اللازم لتقدير القيمة الإجمالية المجتمع باستخدام ($N\,\overline{x}$) بخطأ تقدير (β) فهي :

$$m = \frac{M\,\sigma_{cl}^2}{M\,D + \sigma_{cl}^2} \qquad \text{.... (8 - 20)}$$

حيث $\left(D = \frac{B^2}{Z^2\,M^2}\right)$ ويمكن استخدام (s_{cl}^2) كتقدير لـ (σ_{cl}^2) .

ويمكننا استخدام الصيغة التالية لتحديد حجم العينة اللازم لتقدير القيمة الإجمالية للمجتمع باستخدام ($M\,\overline{x}_{cl}$) أى متوسط قيمة العنقود بخطأ تقدير (β) :

$$m = \frac{M\,\sigma_{cl}^2}{M\,D + \sigma_{cl}^2} \qquad \text{.... (8 - 21)}$$

حيث $\left(D = \frac{B^2}{Z^2\,M^2}\right)$ و (σ_{cl}^2) يمكن تقديره باستخدام (s_{cl}^2) الصيغة (8 - 13) وقيمة (s_{cl}^2) فى هذه الحالة تختلف عن القيمة السابقة لاختلاف الصيغة المستخدمة كما ذكرنا سابقًا ، ولاستخراج حجم العينة العنقودية البسيطة لتقدير نسبة المجتمع نستخدم الصيغة (8 - 19) ، كما نستخدم الصيغة (8 - 20) أو (8 - 21) لتقدير القيمة الكلية حيث نقدر (σ_{cl}^2) من العينة باستخدام :

$$s_{cl}^2 = \frac{\sum_{i=1}^{m} (a_i - p\,n_i)^2}{m - 1}$$

تطبيق (٨ – ٥) :

سحبت عينة استطلاعية لتقدير متوسط الراتب الشهرى للموظف فى إحدى الوزارات ، وقد كان عدد الإدارات (العناقيد) فى الوزارة (٢٥) إدارة . اختير عدد من الإدارات كعينة وقدر تباين العناقيد (٨٣٤٠٠٠٠) . ما هو حجم العينة (عدد الإدارات) اللازم لتقدير متوسط المجتمع ومن ثم لتقدير القيمة الكلية إذا كان خطأ التقدير (١٥٠) ومتوسط حجم العنقود (٢٠) موظفًا ؟

الحـــل :

لدينا : $M = 25$, $B = 150$, $(s_{cl}^2) = 8340000$, $n = 20$

إن حجم العينة (m) يساوى :

$$m = \frac{M \sigma_{cl}^2}{M D + \sigma_{cl}^2}$$

حيث $D = \frac{B^2 \ \overline{N}^2}{Z^2}$ ونظراً لعدم معرفة متوسط حجم عنقود المجتمع $(\overline{N})$ نستخدم متوسط حجم عنقود العينة $(\overline{n})$ كتقدير له ، وكذلك نستخدم تقدير تباين العنقود (s_{cl}^2) لعدم معرفة (σ_{cl}^2) فيكون قيمة (D) بدرجة ثقة ٩٥٪ :

$$D = \frac{(150)^2 (20)^2}{(1.96)^2} = \frac{9000000}{3.8416} = 2342774$$

فيكون عدد الإدارات أى حجم العينة المطلوب :

$$m = \frac{25 \times 8340000}{(25 \times 2342774) + (8340000)}$$

$$= \frac{208500000}{66909350}$$

$$= 3.11 \approx 3$$

أى يتم اختيار (٣) إدارات كعينة عنقودية يتم حصر رواتب موظفيها حصراً شاملاً .

ولتحديد حجم العينة اللازم لتقدير القيمة الكلية للمجتمع نستخدم الصيغة التالية باستخدام $(\hat{T} = N \bar{x})$ نستخدم الصيغة أعلاه مع تبديل قيمة (D) بافتراض خطأ التقدير للقيمة الكلية بحدود (٧٠٠٠٠) :

$$D = \frac{B^2}{Z^2 M^2} = \frac{(70000)^2}{(1.96)^2 (25)^2}$$

$$= \frac{4900000000}{2401} = 2040816$$

وبالتالى يكون حجم العينة المطلوب :

$$m = \frac{25 \times 8340000}{(25 \times 2040816) + (8340000)} = \frac{208500000}{59360400}$$

$$= 3.5 \approx 4$$

أما فى حالة استخدام الصيغة $\hat{T} = M \overline{x}_{cl}$ لتقدير القيمة الكلية فيتطلب ذلك استخراج قيمة $\overline{x}_{cl}$ كما هو موضح فى الصيغة (6 - 8) .

تطبيق (٨ – ٦) :

استخدمت نتائج التطبيق (٨ – ٤) لدراسة مستوى الخدمات المقدمة للمرضى فى المستشفى نفسه . ما هو تقدير حجم العينة اللازم لتقدير نسبة المرضى الذين يرون أن مستوى الخدمات جيد علماً بأن خطأ التقدير المطلوب (٠,٠٥) وبدرجة ثقة (٩٥٪) وعدد الأقسام (٥٠) قسماً ومتوسط حجم العنقود (٢٠) مريضاً ؟ ثم ما هو حجم العينة اللازم لتقدير إجمالى الذين يرون أن مستوى الخدمات جيد إذا كان خطأ التقدير (٤٠) مريضاً ؟

الحـــل :

لدينا البيانات التالية :

$M = 50$, $B = 0.05$ $Z_{1-\alpha/2} = 1.96$, $\overline{N} = 10$

$N = 550$, $p = 0.8182$

ويكون حجم العينة اللازم لتقدير نسبة المجتمع :

$$m = \frac{M \sigma_{cl}^2}{M D + \sigma_{cl}^2}$$

إن تقدير تباين العناقيد من بيانات التطبيق السابق هو :

$$s_{cl}^2 = \frac{\sum_{i=1}^{m} (a_i - p\ n_i)^2}{m - 1}$$

$$= \frac{410}{22 - 1} = \frac{410}{21}$$

$$= 19.52$$

$$D = \frac{\beta^2 \, \overline{N}^2}{Z^2} = \frac{(0.05)^2 \, (20)^2}{(1.96)^2}$$

$$= \frac{1}{3.8416} = 0.26$$

ويكون :

$$m = \frac{50 \times 19.52}{(50 \times 0.26) + 19.52}$$

$$= \frac{976}{32.52} = 30$$

أى عدد الأقسام اللازم لتقدير نسبة المجتمع هو (٣٠) قسماً ، أما عدد الأقسام اللازم لتقدير إجمالي الذين يرون أن مستوى الخدمات جيد فيساوي الصيغة (21 - 8) أى :

$$m = \frac{M \, \sigma^2_{cl}}{M \, D + \sigma^2_{cl}}$$

حيث :

$$D = \frac{\beta^2}{Z^2 \, M^2} = \frac{(40)^2}{(1.96)^2 (50)^2} = \frac{1600}{9604}$$

$$= 0.167$$

فيكون حجم العينة المطلوب :

$$m = \frac{50 \times 19.52}{(50 \times 0.167) + 19.52} = \frac{976}{27.87}$$

$$= 35$$

أى (٣٥) قسماً .

تطبيق (٨ – ٧) :

ترغب إحدى المؤسسات فى تقدير الإنفاق الشهرى للعاملين فى محلاتها البالغ عددها (١٠٠) محل تجارى ونسبة المتزوجين منهم . وقد اختارت عينة عنقودية بسيطة حجمها (٢٢) محلاً ، وقامت بحصر عدد الذين يعملون فيها حصراً شاملاً ، وكانت النتائج كما يلى :

رقم المحل	عدد العاملين	عدد المتزوجين	الإنفاق بالآلاف	رقم المحل	عدد العاملين	عدد المتزوجين	الإنفاق بالآلاف
١	٨	٣	٢٢	١٢	٩	٣	٢٧
٢	٧	٤	٢١	١٣	٨	٤	٢٤
٣	٦	٥	١٨	١٤	٩	٥	٢٧
٤	٧	٥	٢٠	١٥	٨	٤	٢٥
٥	٨	٤	١٦	١٦	٥	٤	٢٢
٦	٦	٣	١٨	١٧	٨	٥	٢٤
٧	٨	٥	٢٤	١٨	٦	٣	٢٣
٨	٩	٧	٢٧	١٩	٥	٢	٢٤
٩	٦	٤	٢٤	٢٠	٨	٤	٢٧
١٠	٦	٤	٢٤	٢١	٦	٤	٢٢
١١	٧	٣	٢٨	٢٢	٤	٣	١٩

المطلوب :

١ – تقدير متوسط عدد العاملين فى المحل وإجمالى عدد العاملين فـى المؤسسة .

٢ – تقدير متوسط الإنفاق الشهرى للعامل وإجمالى الإنفاق الشهرى .

٣ – تقدير نسبة المتزوجين فى المؤسسة وإجمالى عددهم بمستـوى ثقة (٩٥٪) .

الحـــل :

لدينا :

$$\sum_{i=1}^{22} n_i = 154 \ , \sum_{i=1}^{22} a_i = 88 \ , \ m = 22 \ , \sum_{i=1}^{22} x_i = 506$$

١ – تقدير متوسط العاملين في كل محل من محلات المؤسسة .

$$\bar{n} = \sum_{i=1}^{m} n_i / m$$

$$= 154 / 22 = 7$$

أي (٧) عمال . ونعلم أن $(\bar{n})$ هو تقدير لـ $(\bar{N})$ لذا يكون تقدير إجمالي عدد الموظفين :

$$N = \sum_{i=1}^{M} N_i = M\,\bar{N} = M\,\bar{n}$$

$$= 100 \times 7 = 700$$

أي (٧٠٠) عامل .

٢ – نعلم أن متوسط الإنفاق الشهري للعامل :

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} n_i}$$

$$= \frac{506}{154} = 3.286$$

أي (٣٢٨٦) ريالاً .

ولتقدير حدي الثقة نستخدم الصيغة التالية :

$$\bar{x} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(\bar{x})}$$

حيث :

$$\hat{V}(\bar{x}) = \frac{(M - m)}{M m \bar{N}^2} \times \frac{\sum_{i=1}^{m} (x_i - \bar{x} n_i)^2}{m - 1}$$

$$\sum_{i=1}^{m}(x_i - \bar{x} n_i)^2 = \sum_{i=1}^{m} x_i^2 - 2\,\bar{x} \sum_{i=1}^{m} x_i\, n_i + \bar{x}^2 \sum_{i=1}^{m} n_i^2$$

من بيانات التطبيق نجد أن :

$$\sum x_i^2 = 22^2 + 21^2 + \cdots + 19^2 = 11868$$

$$\sum n_i^2 = 8^2 + 7^2 + \cdots + 4^2 = 1120$$

$$\sum x_i\, n_i = (8 \times 22) + (7 \times 21) + \cdots + (4 \times 19) = 3588$$

لذا نجد أن :

$$\sum_{i=1}^{m}(x_i - \bar{x} n_i)^2 = 11868 - (2 \times 3.286 \times 3588) + (3.286)^2\ (1120)$$

$$= 11868 - 23580 + 12094$$

$$= 382$$

ويكون تباين العينة العنقودية البسيطة :

$$s^2 = \frac{\sum_{i=1}^{m} (x_i - \bar{x} n_i)^2}{m - 1}$$

$$= \frac{382}{22 - 1} = 18.19$$

أما تباين تقدير متوسط المجتمع المقدر من بيانات عينة عنقودية بسيطة فيساوى :

$$\hat{V}(\bar{x}) = \frac{100 - 22}{100 \times 22 - (7)^2} \times 18.19$$

حيث :

$$\overline{N} = \overline{n} = 7$$

$$= \frac{78}{2151} \times 18.19 = 0.66$$

ويكون حدا الثقة بدرجة ثقة ٩٥٪ .

$$3.286 \mp 1.96 \sqrt{0.66}$$

$$= 3.286 \mp 1.592$$

أى أن

$$1.694 \le \mu \le 4.878$$

ويمكننا القول إن متوسط الإنفاق للعامل فى المؤسسة يتراوح بين (١٦٩٤) ريالاً و(٤٨٧٨) ريالاً وذلك بدرجة ثقة ٩٥٪ .

أما حدا الثقة لتقدير القيمة الكلية للإنفاق الشهرى فهما :

$$\hat{T} \mp Z_{(1-\alpha/2)} \hat{V}(\hat{T})$$

حيث :

$$\hat{V}(\hat{T}) = \hat{V}(N\bar{x}) = N^2 \hat{V}(\bar{x})$$

$$\hat{T} = N\bar{x} = 700 \times 3.286 = 2300.2$$

$$\hat{V}(\hat{T}) = 700^2 \times 0.66 = 323400$$

وبالتالى يكون حدا الثقة :

$$2300.2 \mp 1.96 \times \sqrt{323400}$$

$$= 2300.2 \mp 1114.618$$

أى أن :

$$1185.582 \le T \le 3414.818$$

أى أن إجمالى الإنفاق الشهرى لمنسوبى المؤسسة يتراوح بين (١١٨٥٥٨٢) ريالاً و(٣٤١٤٨١٨) ريالاً بدرجة ثقة ٩٥٪ .

تطبيق (٨ – ٨) :

باستخدام بيانات التطبيق (٨ – ٧) ما هو تقدير نسبة المتزوجين وتقدير إجمالى عددهم بمستوى ثقة (٩٥٪) ؟

الحـــل :

إذا رمزنا لعدد المتزوجين في العنقود (i) بالرمز a_i يكون لدينا :

$$p = \frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{m} n_i}$$

$$\hat{V}(p) = \frac{(M - m)}{M m \overline{N}^2} \frac{\sum_{i=1}^{m} (a_i - p n_i)^2}{m - 1}$$

من بيانات التطبيق (٨ – ٧) نجد أن :

$$\sum_{i=1}^{22} a_i = 3 + 4 + ----- + 3 = 88$$

$$\sum a_i^2 = 3^2 + 4^2 + ----- + 3^2 = 376$$

$$\sum a_i \; n_i = (8 \text{ x } 3) + (7 \text{ x } 4) + ---- +(4 \text{ x } 3) = 631$$

وبالتالى يكون تقدير نسبة المتزوجين فى المؤسسة :

$$p = \frac{88}{154} = 0.5714$$

أى (٥٧,١٤) ٪ من إجمالى منسوبى المؤسسة .

لاستخراج تقدير التباين لنسبة المتزوجين نجد أن :

$$\sum_{i=1}^{m} (a_i - p\ n_i)^2 = \sum_{i=1}^{m} a_i^2 - 2\ p \sum_{i=1}^{m} a_i\ n_i + p^2 \sum n_i^2$$

$$= 376 - 2 \times 0.5714 \times 631 + (0.5714)^2 \times 1120$$

$$= 376 - 721.1068 + 365.678$$

$$= 20.577$$

$$\hat{V}(p) = \frac{100 - 22}{100 \times 22 \times 7^2} \times \frac{20.577}{22 - 1} = \frac{1605}{2263800}$$

$$= 0.00709$$

وبالتالى نجد حدى الثقة هما :

$$p \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(p)}$$

$$= 0.5714 \mp 1.96 \sqrt{0.0709}$$

$$= 0.5714 \mp 0.165$$

أى أن :

$$0.4060 \le P \le 0.7364$$

ويمكننا القول إنه بمستوى ثقة ٩٥٪ فإن نسبة المتزوجين فى المؤسسة يتراوح بين (٤٠,٦٤٪) و(٧٣,٦٤٪) .

أما تقدير إجمالى عدد المتزوجين ($\hat{T}_A$) فينتج من ضرب هاتين النسبتين بحجم المجتمع (N) أى (٧٠٠) فيكون حدا الثقة لدينا :

$$0.4064 \times 700 \le T_A \le 0.7364 \times 700$$

أى :

$$284 \le T_A \le 515$$

ويمكننا القول إن إجمالى عدد المتزوجين فى المؤسسة يتراوح بين (٢٨٤) متزوجاً و(٥١٥) متزوجاً وذلك بدرجة ثقة ٩٥٪ .

الفصل
التاسع

المعاينة العنقودية ذات المرحلتين
وذات المراحل المتعددة

(Two - Stages and Multi - Stages Cluster Sampling)

٩ - ١ تمهيد :

تستخدم المعاينة العنقودية ذات المرحلتين والمعاينة العنقودية ذات المراحل المتعددة بشكل واسع فى الحياة العملية عندما يكون حجم المجتمع كبيراً ولا يتوافر إطار شامل وحديث بأسماء الوحدات الإحصائية وعناوينها .

نجد فى كثير من الحالات أن المجتمع يتكون من مجموعات (عناقيد) رئيسية تسمى العناقيد الأولية (أو الابتدائية) وكل عنقود يتكون من عدد كبير من الوحدات الإحصائية ولا يتوافر لدينا قائمة بأسماء وعناوين هذه الوحدات التى تشكل وحدات المجتمع ، أو أن إعداد هذا الإطار يتطلب وقتًا طويلاً وإمكانات مادية وبشرية ضخمة لا يمكن توفيرها فى بعض الحالات . يمكننا فى هذه الحالة اختيار عينة عشوائية بسيطة من العناقيد (كمرحلة أولى) . ثم نقوم بإعداد إطار للعناقيد المختارة فقط ، ونختار عينة عشوائية بسيطة من الوحدات من كل عنقود من العناقيد المختارة (كمرحلة ثانية) ونحصرها حصراً شاملاً وبذلك نحصل على المعاينة العنقودية ذات المرحلتين .

وإذا اعتبرنا الوحدات المختارة فى المرحلة الثانية كعناقيد جديدة يتكون كل منها من عدد من الوحدات ، فإننا نختار عدداً من الوحدات من كل عنقود من هذه العناقيد الجديدة ونحصرها حصراً شاملاً ونحصل بذلك على المعاينة ذات المراحل الثلاث (وتسمى المعاينة العنقودية ذات المراحل المتعددة عندما يتم اختيار الوحدات فى ثلاث مراحل أو أكثر) .

وسنقوم بدراسة النوعين التاليين من المعاينات :

- المعاينة العنقودية ذات المرحلتين .

- المعاينة العنقودية ذات المراحل المتعددة .

٩ - ٢ المعاينة العنقودية ذات المرحلتين :

٩ - ٢ - ١ تعريف العينة العنقودية ذات المرحلتين :

يمكننا تعريف العينة العنقودية ذات المرحلتين بأنها "العينة التى نحصل عليها باختيار عينة عشوائية بسيطة من العناقيد كمرحلة أولى ومن ثم اختيار عينة عشوائية بسيطة من الوحدات من كل عنقود من العناقيد المختارة فى المرحلة الأولى (عناقيد العينة) كمرحلة ثانية وحصر العناقيد المختارة فى المرحلة الثانية حصراً شاملاً .

٩ - ٢ - ٢ طريقة اختيار العينة العنقودية ذات المرحلتين :

إذا كان لدينا مجتمع إحصائى مؤلف من (M) عنقوداً أوليًا (مجموعات أولية) ، ورمزنا إلى حجم العنقود (i) فى المجتمع بالرمز (N_i) حيث ($i = 1,2, -----, M$) ونريد اختيار عينة عنقودية ذات مرحلتين من هذا المجتمع . نختار عدداً من العناقيد بطريقة السحب العشوائى وليكن عدد العناقيد المختارة (m) عنقوداً ونقوم بإعداد إطار للوحدات التى يتكون منها كل عنقود من العناقيد المختارة ثم نقوم باختيار عدد من الوحدات بطريقة السحب العشوائى وليكن (n_1) عدد الوحدات المختارة من العنقود الأول من عناقيد العينة و(n_2) هو عدد الوحدات المختارة من العنقود الثانى و(n_i) عدد الوحدات المختارة من العنقود (i) و(n_m) عدد الوحدات المختارة من العنقود الأخير فيكون لدينا عينة عنقودية حجمها (n) مكونة من عدة عينات جزئية أى أن : $n = n_1 + n_2 + ------ + n_m$.

ثم نقوم بحصر العناقيد المختارة حصراً شاملاً .

ولنوضح طريقة اختيار العينة العنقودية ذات المرحلتين بالتطبيق التالى :

تطبيق (٩ - ١) :

نريد اختيار عينة من الأسر من أحد الأحياء لدراسة أحوالهم الاقتصادية والاجتماعية من حيث مستوى الدخل والإنفاق ومتوسط حجم الأسرة وتوزيعاتهم حسب الحالة الزواجية ، ولا يتوافر إطار المساكن لهذا الحى ولا يمكن إعداده لعدم توافر التكاليف المادية والبشرية المطلوبة . فى هذه الحالة ، يمكننا استخدام المعاينة العنقودية ذات المرحلتين وذلك بإجراء الخطوات التالية :

- نعلم أن الحى مقسم إلى عدد من القطاعات وليكن عددها (M) قطاعًا (عنقوداً) يتكون كل منها من عدد من الوحدات ونقوم باختيار عدد من العناقيد ، من عناقيد المجتمع (القطاعات) . وليكن عدد قطاعات المجتمع ($M = 15$) ، اخترنا منها ، ثلاثة قطاعات أى ($m = 3$) ولنفترض أن العناقيد المختارة هى العنقود الثانى والعنقود الثامن والعنقود الثانى عشر .
- نقوم بإعداد إطار يتضمن أسماء رؤساء الأسر ، وأهم المعلومات والبيانات الأخرى ، وقد تبين أن عدد الأسر فى العناقيد المختارة الثلاثة كانت كما يلى :

$N_2 = 600 , N_8 = 800, N_{12} = 300$

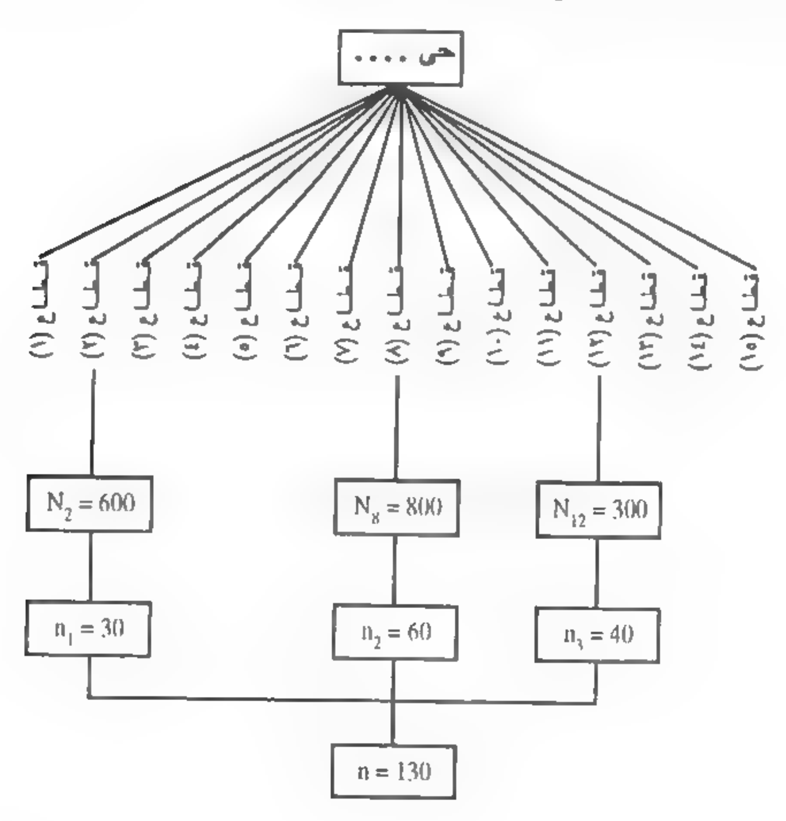
- يتم اختيار عدد من الوحدات (الأسر) من كل عنقود من هذه العناقيد الثلاثة باستخدام إحدى طرق السحب العشوائى ، ولنفترض أن حجم العينات الجزئية كانت كما يلى :
- الوحدات المسحوبة من العنقود الثانى (٣٠) أسرة أى $n_1 = 30$.

- الوحدات المسحوبة من العنقود الثامن (٦٠) أسرة أى $n_2 = 60$.
- الوحدات المسحوبة من العنقود الثانى عشر (٤٠) أسرة أى $n_3 = 40$.

وإجمالى حجم العينة يساوى (١٣٠) أسرة أى :
$$n = n_1 + n_2 + n_3$$
$$= 30 + 60 + 40 = 130$$

- يقوم الباحث بزيارة الأسر المختارة وملء الاستمارات بأجوبة رؤساء الأسر أو ترسل الاستبانات إليهم ليقوموا بملئها بأنفسهم ، ثم نقوم بحصر الأسر المختارة فى المرحلة الثانية حصراً شاملاً .

وهكذا نلاحظ أننا قمنا باختيار عينة عشوائية بسيطة حجمها (m) عنقوداً من عناقيد المجتمع البالغ عددها (M) عنقوداً ، أى اخترنا عينة مكونة من (٣) قطاعات من قطاعات المجتمع المكونة من (١٥) قطاعاً . ثم قمنا باختيار عينة عشوائية بسيطة من الأسر وذلك من كل قطاع من قطاعات العينة أى اخترنا (n_1 , n_2 , n_3) . لذا سميت هذه العينة بالعينة العنقودية ذات المرحلتين ولنوضح ما سبق بالرسم التالى :



وسيتم التطرق إلى كيفية تحديد عدد عناقيد العينة وعدد وحدات العينة عند دراستنا لكيفية تحديد حجم العينة في الصفحات القادمة .

٩ - ٢ - ٣ تقديرات أهم معالم المجتمع :

أ - تقدير الوسط الحسابي للمجتمع والقيمة الكلية للمجتمع :

ذكرنا فيما سبق أننا اخترنا (m) عنقوداً من عناقيد المجتمع البالغ عددها (M) عنقوداً ، ولنسمى عناقيد العينة المختارة بالعناقيد النهائية وعناقيد المجتمع بالعناقيد الأولية . وهكذا نجد أن لدينا (m) عنقوداً نهائياً .

إن مجموع مفردات العنقود النهائي الأول (x_1) يساوى :

$$x_1 = x_{11} + x_{12} + \text{-----} + x_{1n_1} = \sum_{j=1}^{n_1} x_{1j}$$

ومجموع مفردات العنقود الثاني يساوى .

$$x_2 = x_{21} + x_{22} + \text{-----} + x_{2n_2} = \sum_{j=1}^{n_2} x_{2j}$$

ومجموع مفردات العنقود (i) يساوى .

$$x_i = x_{i1} + x_{i2} + \text{-----} + x_{in_i} = \sum_{j=1}^{n} x_{ij}$$

ومجموع مفردات العنقود النهائي الأخير يساوى :

$$x_m = x_{m1} + x_{m2} + \text{-----} + x_{mn_m} = \sum_{j=1}^{n_m} x_{mj}$$

ومجموع مفردات عناقيد العينة النهائية ولنرمز له بالرمز (x) يساوى :

$$x = x_1 + x_2 + \text{-----} + x_m = \sum_{i=1}^{m} x_i$$

إن متوسط العنقود النهائي (i) ولنرمز له بالرمز $\bar{\bar{x}}_i$ (حيث وضعنا الرمز = فوق x_i للدلالة على أن المقدر للعينة ذات المرحلتين أى أن اختيار الوحدات قد تم فى المرحلة الثانية) يساوى :

$$\bar{\bar{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (9 - 1)}$$

وبالتالي فإن مقدر مجموع قيم العنقود الأول الذى عدد مفرداته (N_1) مفردة يساوى :

$$\hat{X}_1 = N_1 \ \bar{\bar{x}}_1$$

ومقدر مجموع قيم العنقود (i) الذى عدد مفرداته (N_i) هو :

$$\hat{X}_i = N_i \ \bar{\bar{x}}_i \qquad \text{.... (9 - 2)}$$

ولدينا (m) مقدراً أى $(i = 1 , 2 , ----- , m)$ ويكون مجموع تقديرات العناقيد النهائية (عناقيد العينة) مساوياً لـ :

$$\hat{X} = \hat{X}_1 + \hat{X}_2 + ---- + \hat{X}_m$$

أى :

$$\hat{X} = \sum_{i=1}^{m} \hat{X}_i$$

$$\hat{X} = \sum_{i=1}^{m} N_i \ \bar{\bar{x}}_i$$

وبتبديل $\bar{\bar{x}}_i$ بقيمتها من الصيغة (9 - 1) نجد أن :

$$\hat{X} = \sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (9 - 3)}$$

وبافتراض أن حجم العناقيد متساوية تقريباً وعددها (m) عنقوداً نجد أن مقدر متوسط العناقيد ولنرمز له بالرمز ($\hat{\bar{X}}$) يساوى :

$$\hat{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m}\frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (9 - 4)}$$

ويعد هذا المقدر مقدراً غير متحيز لمتوسط العنقود .

ولتقدير مجموع قيم المجتمع باستخدام عينة عنقودية ، نفترض أن حجم جميع عناقيد المجتمع متساوية تقريباً أى أن :

$N_1 = N_2 = \text{------} = N_i = \text{------} = N_M$

وحيث لدينا (M) عنقوداً تمثل عناقيد المجتمع ، لذا فإن مقدر مجموع قيم المجتمع ولنرمز له بالرمز ($\hat{X}$) يساوى مقدر متوسط قيم العنقود للمجتمع (الصيغة (4 - 9)) مضروباً فى عددها (M) عنقوداً أى أن $\hat{X} = M\,\hat{\bar{X}}$.

ونجد أن :

$$\hat{X} = \frac{M}{m}\sum_{i=1}^{m}\frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (9 - 5)}$$

ويكون مقدر متوسط المجتمع (مقدر متوسط المفردة) مساوياً لـ :

$$\hat{\bar{\bar{X}}} = \hat{\mu} = \frac{1}{N}\hat{X}$$

أى أن :

$$\hat{\bar{\bar{X}}} = \frac{M}{Nm}\sum_{i=1}^{m}\frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij} \qquad \text{.... (9 - 6)}$$

حيث $N = M\,\bar{N}$ ويتم حساب متوسط حجم عنقود المجتمع من أحجام العناقيد المختارة أى أن :

$$\bar{N} = \frac{1}{m}\sum_{i=1}^{m} N_i$$

ويعد هذا المقدر $\hat{\bar{X}}$ مقدراً غير متحيز لمتوسط المجتمع ، أما متوسط العينة العشوائية البسيطة للوحدات ولنرمز له ($\bar{x}_{ran}$) فيساوى :

$$\bar{x}_{ran} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad .... (9 - 7)$$

وطبعًا يختلف هذا المتوسط عن المتوسط المقدر بالصيغة (6 - 9) أى متوسط العينة العنقودية .

ولابد لنا من الإشارة فى هذا المجال إلى أن عدد العينات الممكن سحبها للعناقيد الابتدائية هو $\binom{M}{m}$، أما عدد العينات الممكن سحبها للوحدات أى لـ $n_1 , n_2 , ----- , n_m$ فيساوى :

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times ---- \times \binom{N_m}{n_m}$$

ولابد لنا من الإشارة إلى أنه فى حالة عدم معرفة حجم المجتمع (N) يتم تقديره عن طريق تقدير متوسط حجم العنقود من بيانات العينة وضربه فى عدد العناقيد أى :

$$N = M \sum_{i=1}^{m} \frac{N_i}{m}$$

تطبيق (٩ – ٢) :

تتكون إحدى المناطق من (١٠) حيازات زراعية يبلغ متوسط عدد العاملين فى كل منها (٤٠) عاملاً تقريبًا ، ونريد سحب عينة من حيازتين ، وذلك لتقدير متوسط الراتب الذى يتقاضاه العامل شهريًا فى الحيازة ، وإجمالى الرواتب التى يتقاضاها عمال الحيازات ، وقد كانت رواتب العمال للعينة التى تم اختيارها عشوائيًا (بالآلاف) كما يلى :

$$x_{11} = 3, x_{12} = 5, x_{13} = 4, x_{14} = 5, x_{15} = 4, x_{16} = 3$$

$$x_{21} = 2, x_{22} = 3, x_{23} = 3, x_{24} = 4$$

المطلوب :

– تقدير متوسط الراتب الشهرى الذى يتقاضاه العامل فى الحيازة .
– تقدير إجمالى الرواتب الشهرية التى يتقاضاها العمال فى الحيازات .

الحـــل :

من بيانات التطبيق ، لدينا البيانات التالية :

$N = 40 \times 10 = 400$, $n = n_1 + n_2 = 6 + 4 = 10$
$M = 10$, $m = 2$, $n_1 = 6$, $n_2 = 4$
$N_1 = 40$, $N_2 = 40$

- متوسط العنقود الأول والعنقود الثانى :

لاستخراج متوسط العنقود (i) نستخدم (الصيغة (1 - 9)) :

$$\bar{\bar{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{x_i}{n_i}$$

حيث :

$$x_i = \sum_{j=1}^{n_i} x_{ij}$$

وبذلك يكون متوسط العنقود الأول من بيانات العينة :

$$\bar{\bar{x}}_1 = \frac{x_1}{n_1}$$

$$= \frac{3 + 5 + ----- + 3}{6} = \frac{24}{6} = 4$$

$$\bar{\bar{x}}_2 = \frac{x_2}{n_2}$$

$$= \frac{2 + 3 + 3 + 4}{4} = \frac{12}{4} = 3$$

– مجموع قيم العنقودين :

$$x = \sum_{i=1}^{m} x_i$$

$$= x_1 + x_2$$

$$= 24 + 12 = 36$$

– إن تقدير مجموع قيم العنقود (i) يساوى (الصيغة 2 - 9) .

$$\hat{X}_i = N_i \bar{\bar{x}}_i$$

وبذلك يكون تقدير مجموع العنقود الأول :

$$\hat{X}_1 = N_1 \bar{\bar{x}}_1$$

$$= 40 \times 4 = 160$$

ويكون تقدير مجموع العنقود الثانى :

$$\hat{X}_2 = N_2 \bar{\bar{x}}_2$$

$$= 40 \times 3 = 120$$

ويكون تقدير مجموع العنقودين :

$$\hat{X} = \sum_{i=1}^{m} N_i \bar{\bar{x}}_i$$

$$= 160 + 120 = 280$$

وباستخدام الصيغة (3 - 9) نجد أن هذا التقدير يساوى :

$$\hat{X} = \sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$= (\frac{40}{6} \times 24) + (\frac{40}{4} \times 12)$$

$$= 160 + 120 = 280$$

وهو الجواب السابق نفسه .

ولاستخراج تقدير متوسط العناقيد نستخدم الصيغة (4 - 9) :

$$\hat{\bar{X}} = \frac{1}{m} \sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{1}{2} \left[(\frac{40}{6} \text{ x } 24) + (\frac{40}{4} \text{ x } 12) \right]$$

$$= \frac{1}{2} (160 + 120)$$

$$= \frac{280}{2} = 140$$

أى يساوى : $\frac{\hat{X}}{m}$

– لاستخراج تقدير متوسط الراتب الذى يتقاضاه العامل ، نستخدم الصيغة (6 - 9) حيث نضرب تقدير متوسط العنقود ($\hat{\bar{X}}$) فى عدد عناقيد المجتمع (M) ونقسم الناتج على حجم المجتمع أى :

$$\hat{\bar{X}} = \frac{M}{Nm} \sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{10}{400 \text{ x } 2} \left[(\frac{40}{6} \text{ x } 24) + (\frac{40}{4} \text{ x } 12) \right]$$

$$= \frac{10}{800} [(160 + 120)]$$

$$= \frac{280}{80} = 3.5$$

أى أن تقدير متوسط الراتب الذى يتقاضاه عامل الحيازات هو (٣٥٠٠) ريال .

– أما تقدير إجمالى الرواتب فهو عبارة عن المتوسط مضروباً بحجم المجتمع أى :

$$\hat{T} = N \hat{\mu}$$

$$= 400 \text{ x } 3.5 = 1400$$

ويمكن الحصول على الجواب نفسه مباشرة باستخدام الصيغة :

$$\hat{T} = M \hat{\bar{X}}$$

$$= 10 \text{ x } 140 = 1400$$

أى أن تقدير إجمالى الرواتب الشهرية لعمال الحيازات يبلغ (١٤٠٠٠٠) .

– أما تقدير متوسط المجتمع باستخدام المعاينة العشوائية البسيطة فيتم استخراجه كما يلى :

$$\bar{x}_{ran} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{24 + 12}{10} = 3.6$$

وتقدير مجموع المجتمع :

$$\widehat{X} = N\,\bar{x}$$

$$= 400 \times 3.6 = 1440$$

ويختلف هذان التقديران عن تقديرى المعاينة العنقودية اللذين حصلنا عليهما فيما سبق .

ب – تباين تقدير القيمة الكلية وتقديره :

إن سحب وحدات المعاينة العنقودية ذات المرحلتين يتم على مرحلتين :

– سحب (m) وحدة معاينة ابتدائية من (M) وحدة ابتدائية (m عنقوداً) (Prunary Sampling Units) .

– سحب (n_i) وحدة معاينة ثانوية من وحدات كل عنقود (N_i) حيث ($i = 1 , 2 , ----- , m$) أى سحب ($n_1 , n_2 , ----- , n_m$) من ($N_1 , N_2 , ----- , N_m$) على التوالى وتسمى (Secondary Sampling Units) لذا عند استخراج تباين تقدير القيمة الكلية ولنرمز له بالرمز $V(\widehat{X})$ لابد من التمييز بين تباين وحدات المعاينة الابتدائية والتباين داخل وحدات المعاينة الابتدائية ، وهكذا يجب التفريق بين التباينين التاليين :

– التباين بين وحدات المعاينة الابتدائية .

– التباين داخل وحدات المعاينة الابتدائية ، ونجد أن تباين تقدير القيمة الكلية هو عبارة عن حاصل جمع هذين التباينين ، أى أن :

تباين $(\widehat{X})$ = التباين بين الوحدات + داخل الوحدات والصيغة المستخدمة لاستخراج قيمة هذا التباين تساوى :

$$V(\widehat{X}) = \left(\frac{M^2}{m}\,\frac{M - m}{M}\,S_b^2\right) + \left(\frac{M}{m}\sum_{i=1}^{M} N_i^2\,\frac{N_i - n_i}{N_i}\,\frac{S_i^2}{n_i}\right) \qquad .... (9 - 8)$$

$$S_b^2 = \frac{1}{M-1} \sum_{i=1}^{M} (X_i - \overline{X}) \qquad \text{.... (9 - 9)}$$

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (X_{ij} - \overline{\overline{X}}_i)^2 \qquad \text{.... (9 - 10)}$$

حيث (S_b^2) هو التباين بين العناقيد الابتدائية و(S_i^2) يرمز إلى التباين داخل العناقيد .

كما أن $(\overline{\overline{X}}_i)$ هو متوسط قيمة الوحدة في العنقود أي :

$$\overline{\overline{X}}_i = \frac{X_i}{N_i}$$

ومتوسط قيمة العنقود

$$\overline{X} = \frac{X}{M}$$

وفي التطبيقات العملية ، خاصة عندما يكون حجم المجتمع كبيرًا ، نجد أن هذا التباين يكون مجهولاً ويتم تقديره من بيانات العينة .

إن مقدر تباين تقدير القيمة الكلية ولنرمز له بالرمز $\hat{V}(\hat{X})$ يساوي :

$$\hat{V}(\hat{X}) = \left(\frac{M^2}{m} \frac{M-m}{M} s_b^2\right) + \left(\frac{M}{m} \sum_{i=1}^{m} N_i^2 \frac{N_i - n_i}{N_i} \frac{s_i^2}{n_i}\right) \qquad \text{.... (9 - 11)}$$

حيث :

$$s_b^2 = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{X}_i - \hat{\overline{X}})^2 \qquad \text{.... (9 - 12)}$$

$$s_i^2 = \frac{1}{m_i - 1} \sum_{i=1}^{n_i} (x_{ij} - \overline{\overline{x}}_i)^2 \qquad \text{.... (9 - 13)}$$

كما أن :

$$\hat{X}_i = N_i \bar{\bar{x}}_i , \bar{\bar{x}}_i = \frac{x_i}{n_i} , \hat{\bar{X}} = \frac{1}{m} \sum_{i=1}^{m} \hat{X}_i$$

ونلاحظ أن (s_i^2) يظهر تباين (x_{ij}) داخل الوحدات النهائية من وحدات المعاينة الثانوية .

إن $\hat{V}(\hat{X})$ هو مقدر غير متحيز لـ $V(\hat{X})$. كذلك لابد من الإشارة الى أن (s_i^2) هو مقدر غير متحيز لـ (S_i^2) ولكن (s_b^2) هو مقدر متحيز لـ (S_b^2) ومقدار التحيز هو :

$$\frac{1}{M} \sum_{i=1}^{M} N_i^2 \frac{N_i - n_i}{N_i} \frac{S_i^2}{n_i}$$

أما مقدر تباين تقدير متوسط المجتمع $\hat{V}(\hat{\bar{X}})$ فهو عبارة عن :

$$\hat{V}(\hat{\bar{X}}) = \hat{V}\left(\frac{\hat{X}}{N}\right) = \hat{V}(\hat{\mu})$$

$$\hat{V}(\hat{\bar{X}}) = \frac{1}{\bar{N}^2} \hat{V}(\hat{X})$$

حيث نستخدم الصيغة (11 - 9) لاستخراج قيمة $\hat{V}(\hat{\bar{X}})$ و$(\bar{N})$ تمثل متوسط حجم العنقود ونجد أن :

$$\hat{V}(\hat{\bar{X}}) = \frac{1}{\bar{N}^2} \left[\left(\frac{M^2}{m} \frac{M-m}{M} \right) s_b^2 + \left(\frac{M}{m} \sum_{i=1}^{m} N_i^2 \frac{N_i - n_i}{N_i} \frac{s_i^2}{n_i} \right) \right] \quad .... (9 - 14)$$

حيث (s_b^2) و (s_i^2) موضحتان في الصيغتين (12 - 9) و (13 - 9) .

تطبيق (٩ – ٣) :

لدينا ثلاث إدارات (A , B , C) وليكن (X_{ij}) يمثل عدد سنوات الخبرة لدى الموظف (j) في العنقود (الإدارة) (i) . لنَخْتَرْ عشوائيًا باستخدام جداول الأرقام العشوائية

عينة من إدارتين (أى عنقودين $m = 2$) ثم نختار موظفين من كل إدارة من الإدارات لنختر $(n_1 = n_2 = 2)$.

المطلوب :

استخراج

١ – عدد العينات الممكن سحبها وما هى هذه العينات .

٢ – تقدير متوسط سنوات الخبرة للموظف وإجمالى سنوات الخبرة لديهم . علما بأن سنوات الخبرة للموظفين فى الادارات الثلاث كانت كما يلى :

العنقود (A) $X_{13} = 5$, $X_{12} = 3$, $X_{11} = 1$

العنقود (B) $X_{23} = 7$, $X_{22} = 5$, $X_{21} = 3$

العنقود (C) $X_{33} = 9$, $X_{32} = 7$, $X_{31} = 5$

الحـــل :

– إن عدد العينات الممكنة للعناقيد الابتدائية فى المجتمع يساوى :

$$\binom{M}{m} = \binom{3}{2} = \frac{3!}{2!\,1!} = \frac{6}{2} = 3$$

ومجموع عدد العينات الممكنة من $m = 2$ أى لـ (n_1 , n_2) يساوى :

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} = \binom{3}{2} \times \binom{3}{2} = 3 \times 3 = 9$$

ومجموع عدد العينات الممكنة للعينة العنقودية :

$$\binom{M}{m}\binom{N_1}{n_1}\binom{N_2}{n_2} = 3 \times 3 \times 3 = 27$$

ويوضح الجدول التالي هذه العينات :

العينات الممكن سحبها

A	B	$\hat{X}$	A	C	$\hat{X}$	B	C	$\hat{X}$
1,3	3,5	27	1,3	5,7	36	3,5	5,7	45
1,3	3,7	31.5	1,3	5,9	40.5	3,5	5,9	49.5
1,3	5,7	39	1,3	7,9	45	3,5	7,9	54
1,5	3,5	31.5	1,5	5,7	40.5	3,7	5,7	49.5
1,5	3,7	36	1,5	5,9	45	3,7	5,9	54
1,5	5,7	40.5	1,5	7,9	43.5	3,7	7,9	58.5
3,5	3,5	36	3,5	5,7	45	5,7	5,7	54
3,5	3,7	40.5	3,5	5,9	49.5	5,7	5,9	58.5
3,5	5,7	45	3,5	7,9	54	5,7	7,9	63
Total		324			405			486

لنوضح فيما يلي أهم المتوسطات باستخدام بيانات المجتمع ، ومن ثم لنستخدم بيانات العينة الأولى الممكن سحبها لتوضيح كيفية تقدير متوسط المجتمع والقيمة الكلية للمجتمع (مفردات العينة الأولى هي 1,3 من العنقود (A) و (3,5) من العنقود (B) .

- باستخدام بيانات المجتمع نجد أن قيمة العناقيد ومتوسطاتها هي :

$$X_i = \sum_{j=1}^{N_i} X_{ij} \, , \; \bar{\bar{X}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}$$

ويكون قيمة مفردات العناقيد الأول والثاني والثالث ومتوسطاتها هي :

$$X_1 = 1 + 3 + 5 = 9 \quad , \bar{\bar{X}}_1 = 3$$

$$X_2 = 3 + 5 + 7 = 15 \quad , \bar{\bar{X}}_2 = 5$$

$$X_3 = 5 + 7 + 9 - 21 \quad , \bar{\bar{X}}_3 = 7$$

ويكون متوسط العنقود من المجتمع (متوسط سنوات الخبرة للإدارة) :

$$\overline{X}_{cl} = \frac{\sum_{i=1}^{M} X_i}{M} = \frac{9 + 15 + 21}{3} = \frac{45}{3} = 15$$

أما القيمة الكلية.(إجمالى سنوات الخبرة) فتساوى :

$$T = X = M\ \overline{X}cl = 3 \times 15 = 45$$

ويكون متوسط سنوات الخبرة للموظف الواحد باستخدام إحدى الصيغ التالية :

$$\overline{X} = \frac{X}{N} = \frac{M}{N}\sum_{i=1}^{M} X_i = \frac{M}{N}\sum_{i=1}^{M}\sum_{j=1}^{N_i} X_{ij} = \mu$$

أى أن :

$$\mu = \frac{45}{9} = 5$$

وهكذا نجد أن متوسط سنوات الخبرة لدى الموظف هى (٥) سنوات وذلك باستخدام بيانات المجتمع .

- باستخدام بيانات العينة الأولى المسحوبة من العنقودين الأول والثانى أى الإدارتين (A , B) .

$$(x_{11} = 1\ ,\ x_{12} = 3\ ,\ x_{21} = 3\ ,\ x_{23} = 5$$

نجد أن تقدير سنوات الخبرة لدى موظفى الإدارات الثلاث :

$$\hat{T} = \hat{X} = \frac{M}{m}\sum_{i=1}^{m} x_i\ \frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{M}{m}\left[\frac{N_1}{n_1}\sum_{j=1}^{n_1} x_{1j} + \frac{N_2}{n_2}\sum_{j=1}^{n_2} x_{2j}\right]$$

$$= \frac{3}{2}\left[\frac{3}{2}(1+3)+\frac{3}{2}(3+5)\right]$$

$$= \frac{3}{2}(6+12) = 27$$

أى أن تقدير سنوات الخبرة لجميع الإدارات هو (٢٧) سنة .

- إن تقدير سنوات الخبرة للإدارة (i) هو :

$$x_i = \frac{N_i}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

$$= N_i \bar{\bar{x}}_i$$

إن متوسط العنقود (i) من بيانات العينة هو :

$$\bar{\bar{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

لذا نجد أن :

$$\bar{\bar{x}}_1 = \frac{1+3}{2} = \frac{4}{2} = 2$$

$$\bar{\bar{x}}_2 = \frac{3+5}{2} = \frac{8}{2} = 4$$

وبالتالى يكون تقدير القيمة الكلية لكل من العنقودين B , A على التوالى :

$$x_1 = N_1 \bar{\bar{x}}_1$$

$$= 3 \times 2 = 6$$

$$x_2 = N_2 \bar{\bar{x}}_2$$

$$= 3 \times 4 = 12$$

ويكون تقدير متوسط عدد سنوات الخبرة الموظف فى العنقود (الإدارة) :

$$\widehat{\overline{X}}_{cl} = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$= \frac{x_1 + x_2}{2}$$

$$= \frac{6 + 12}{2} = 9$$

وحيث لدينا ثلاث إدارات ، لذا نجد أن سنوات الخبرة لجميع موظفى الإدارات يساوى :

$$\hat{T} = \hat{X} = M\ \widehat{\overline{X}}_{cl}$$

$$= 3 \times 9 = 27$$

أى (٢٧) سنة .

ونلاحظ فى الجدول السابق أننا قدرنا سنوات الخبرة ($\hat{X}$) للعينات الـ (٢٧) الممكن سحبها ، ويعد ($\hat{X}$) تقديراً غير متحيز للقيمة الكلية للمجتمع أى لسنوات الخبرة للموظفين فــى الإدارات الثلاث .

أما تقدير متوسط سنوات الخبرة للموظف الواحد فيساوى :

$$\widehat{\overline{X}} = \hat{\mu} = \frac{\hat{X}}{N}$$

$$= \frac{M}{Nm}\sum_{i=1}^{m} \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$= \frac{27}{9} = 3$$

أى أن تقدير متوسط سنوات الخبرة هو (٣) سنوات ، ويعد هذا التقدير تقديراً غير متحيز لمتوسط المجتمع أى لمتوسط سنوات الخبرة .

تطبيق (٩ – ٤) :

باستخدام بيانات التطبيق (٩ – ٣) ، استخرج :

١ – تباين تقدير القيمة الكلية $V(\hat{X})$.

٢ – تقدير تباين القيمة الكلية المقدرة $\hat{V}(\hat{X})$.

الحــل :

– إن تباين تقدير القيمة الكلية $V(\hat{X})$ يساوى :

$$V(\hat{X}) = \frac{M^2}{m}\,\frac{M-m}{M}\,S_b^2 + \frac{M}{m}\sum_{i=1}^{M} N_i^2\,\frac{N_i - n_i}{N_i}\,\frac{S_i^2}{n_i}$$

إن :

$$S_b^2 = \frac{1}{M-1}\sum_{i=1}^{M}(X_i - \overline{X})^2$$

$$\overline{X}_{cl} = \frac{X}{M} = \frac{X_1 + X_2 + X_3}{3}$$

$$= \frac{9+15+21}{3} = \frac{45}{3} = 15$$

وبالتالى نجد أن :

$$S_b^2 = \frac{1}{3-1}\left[(9-15)^2 + (15-15)^2 + (21-15)^2\right] = 36$$

كذلك نجد أن :

$$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(X_{ij} - \overline{\overline{X}}_i)^2$$

$$S_1^2 = \frac{1}{N_1 - 1}\sum_{j=1}^{N_1}(X_{1j} - \overline{\overline{X}}_1)^2$$

$$= \frac{1}{3-1}\left[(1-3)^2 + (3-3)^2 + (5-3)^2\right] = 4$$

$$S_2^2 = \frac{1}{3-1}\left[(3-5)^2 + (5-5)^2 + (7-5)^2\right] = 4$$

$$S_3^2 = \frac{1}{3-1}\left[(5-7)^2 + (7-7)^2 + (9-7)^2\right] = 4$$

وبالتالي يكون :

$$V(\hat{X}) = \left[\frac{3^2}{2} \times \frac{3-2}{3} \times 36\right] + \frac{3}{2}\left[3^2 \times \frac{(3-2)}{3} \; \frac{(4+4+4)}{2}\right]$$

$$= 54 + 27 = 81$$

- باستخدام بيانات العينة الأولى الممكن سحبها الأولى ، يمكننا استخراج تقدير تباين تقدير القيمة الكلية $\hat{V}(\hat{X})$ من الصيغة التالية :

$$\hat{V}(\hat{X}) = \frac{M^2}{m} \; \frac{M-m}{M} \; s_b^2 + \frac{M}{m} \sum_{i=1}^{m} N_i^2 \; \frac{N_i - n_i}{N_i} \; \frac{s_i^2}{n_i}$$

حيث :

$$s_b^2 = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{X}_i - \hat{\bar{X}})^2$$

$$\hat{X}_i = N_i \bar{\bar{x}}_i \, , \; \hat{\bar{X}} = \frac{1}{m} \sum_{i=1}^{m} \hat{X}_i$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}}_i)^2$$

لذا نجد أن :

$$\hat{X}_1 = x_1 = N_1 \bar{\bar{x}}_1$$

$$= 3 \; \frac{(1+3)}{2} = 6$$

$$\widehat{X}_2 = N_2 \bar{\bar{x}}_2$$

$$= 3 \frac{(3+5)}{2} = 12$$

$$\widehat{\widehat{X}} = \frac{1}{2}(6+12) = 9$$

وحيث لدينا $\bar{\bar{x}}_1 = 2 , \bar{\bar{x}}_2 = 4$ لذا نجد أن :

$$s_b^2 = \frac{1}{2-1}\left[(6-9)^2 + (12-9)^2\right] = 18$$

$$s_1^2 = \frac{1}{2-1}\left[(1-2)^2 + (3-2)^2\right] = 2$$

$$s_2^2 = \frac{1}{2-1}\left[(3-4)^2 + (5-4)^2\right] = 2$$

$$\widehat{V}(\widehat{X}) = \left[\frac{3^2}{2} \times \frac{3-2}{3} \times 18\right] + \left[\frac{3}{2} \times 3^2 \times \frac{3-2}{3} \times \frac{1}{2}(2+2)\right]$$

$$= 27 + 9 = 36$$

٩ – ٢ – ٤ حدود الثقة لتقدير القيمة الكلية وتقدير متوسط المجتمع :

يمكننا استخراج حدى الثقة لتقدير القيمة الكلية للمجتمع ، باستخدام الصيغة التالية :

$$\widehat{X} \pm t_{(1-\alpha/2,n-1)} \sqrt{\widehat{V}(\widehat{X})} \qquad \text{.... (9 -15)}$$

كذلك نستخدم الصيغة التالية لاستخراج حدى الثقة لتقدير متوسط المجتمع :

$$\widehat{\widehat{X}} \pm t_{(1-\alpha/2,n-1)} \sqrt{\widehat{V}(\widehat{\widehat{X}})} \qquad \text{.... (9 -16)}$$

حيث :

$t_{1-\alpha/2,n-1}$ القيمة الجدولية من توزيع (t) بمستوى ثقة %(1 - α) درجات حرية (n - 1).

$\hat{V}(\hat{X})$ تقدير تباين تقدير القيمة الكلية للمجتمع (الصيغة ٩ – ١١)

$\hat{V}(\hat{\bar{X}})$ تقدير تباين تقدير متوسط المجتمع وتساوى $\hat{V}(\hat{\mu})$ (الصيغة (14 - 9)) .

ونستخدم Z (القيمة الجدولية للتوزيع الطبيعى) عندما يكون حجم العينة كبيرًا (٣٠ فاكثر) .

تطبيق (٩ – ٥) :

باستخدام بيانات التطبيقين (٩ – ٣) و (٩ – ٤) ، أوجد حدى الثقة لتقدير إجمالى سنوات الخبرة للموظفين .

الحـل :

إن حدى الثقة لتقدير القيمة الكلية (تقدير إجمالى سنوات الخبرة للموظفين) يساوى :

$$\hat{X} \mp t_{(1-\alpha/2,n-1)} \sqrt{\hat{V}(\hat{X})}$$

وباستخدام نتائج التطبيقين (٩ – ٣) و (٩ – ٤) نجد أن هذين الحدين بمستوى ثقة (٪٩٥) :

$$27 \pm 3.182 \sqrt{36}$$

حيث :

$$t_{(1-\alpha/2,n-1)} = t_{(1-0.05/2,4-1)} = 3.182$$

$$= 27 \mp 19.09$$

ويكون الحد الأدنى :

$$27 - 19.09 = 7.91$$

والحد الأعلى :

$$27 + 19.09 = 46.09$$

أى أنه بدرجة ثقة ٩٥٪ ، فإن إجمالى سنوات الخبرة ستقع بين 7.91 و 46.09 سنة أى

$7.91 \leq X \leq 46.09$.

تطبيق (٩ – ٦) :

ترغب إحدى المؤسسات فى تقدير متوسط راتب العامل الشهرى وتقدير إجمالى رواتب منسوبيها الذين يعملون فى (٩٠) مشروعًا موزعة فى جميع مناطق المملكة ، واستخدمت العينة العنقودية ذات المرحلتين ، حيث تم اختيار عشرة مشاريع ثم اختير حوالى ٢٠٪ من العاملين فى كل مشروع علمًا بأن عدد العاملين فى المؤسسة هو (٤٥٠٠) عامل .

إذا كانت لدينا البيانات التالية (الرواتب بالآلاف) :

$$M = 90 \;,\; m = 10 \;,\; \widehat{\bar{X}} = 4.80 \;,\; \hat{V}(\hat{X}) = 92$$

$$N = 4500 \;,\; \hat{X} = 21600$$

أوجد حدى الثقة لتقدير إجمالى الرواتب وحدى الثقة لتقدير متوسط الراتب للعامل .

الحـل :

إن حدى الثقة لتقدير إجمالى الرواتب يساوى :

$$\hat{X} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{X})}$$

$$= 21600 \pm 1.96\sqrt{92}$$

$$= 21600 \pm 18.8$$

ويكون الحد الأدنى :

21581.2

والحد الأعلى :

21618.8



أى أن إجمالى الرواتب الشهرية لمنسوبى المؤسسة بدرجة ثقة ٩٥٪ يتراوح بين (٢١٥٨١,٢) ألف ريال و (٢١٦١٨,٨) ألف ريال أى :

$$21581.2 \leq x \leq 21618.8$$

أما حدا الثقة لتقدير متوسط الراتب الشهرى فيساوى :

$$\hat{\bar{X}} \pm Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{\bar{X}})}$$

إن :

$$\hat{V}(\hat{\bar{X}}) = \frac{1}{\bar{N}^2} \hat{V}(\hat{X})$$

$$\bar{N} = \frac{4500}{90} = 50$$

$$\hat{V}(\hat{\bar{X}}) = \frac{1}{50^2} \times 92 = 0.037$$

وبالتالى يكون حدا الثقة بدرجة ثقة ٩٥٪ $(\alpha = 0.05)$.

$$4.8 \mp 1.96 \sqrt{0.037}$$

$$= 4.8 \mp 0.38$$

ويكون الحد الأدنى :

$$4.8 - 0.38 = 4.42$$

والحد الأعلى :

$$4.8 + 0.38 = 5.18$$

أى أن متوسط الراتب الشهرى للعامل فى المؤسسة يتراوح بدرجة ثقة ٩٥٪ بين (٤٤٢٠) ريالاً و (٥١٨٠) ريالاً أى :

$$4420 \leq \mu \leq 5180$$

٩ - ٢ - ٥ تقدير نسبة المجتمع :

(Estimation of population proportion)

كثيراً ما نرغب فى تقدير نسبة الذين يتصفون بخاصية معينة باستخدام المعاينة العنقودية ذات المرحلتين . مثلاً قد نرغب فى تقدير نسبة الموافقين على مرشح معين أو إجراء معين . نستخدم فى هذه الحالة الصيغة التالية لتقدير نسبة المجتمع $(\hat{P} = p)$ الذى يعد مقدراً غير متحيز لنسبة المجتمع :

$$\hat{P} = \frac{\sum_{i=1}^{m} N_i \, p_i}{\sum_{i=1}^{m} N_i} \qquad \text{.... (9 -17)}$$

حيث (p_i) تمثل مفردات الذين يتصفون بالظاهرة فى العنقود (i) الذى تم اختياره بالعينة (بافتراض أن عدد الموافقين على إجراء ما فى العنقود (i) من الأشخاص الذين تم اختيارهم (n_i) يساوى (a_i) فإن :

$$p_i = \frac{a_i}{n_i}$$

أما الصيغة المستخدمة لتقدير تباين تقدير نسبة المجتمع فهى :

$$\hat{V}(\hat{P}) = \left[\frac{M - m}{M} \; \frac{1}{m \, \overline{N}^2} \; s_b^2\right] + \left[\frac{1}{m \, M \, \overline{N}^2} \sum_{i=1}^{m} N_i^2 \; \frac{N_i - n_i}{N_i} \; \frac{p_i \, q_i}{n_i - 1}\right] \qquad \text{.... (9 -18)}$$

حيث :

$$s_b^2 = \frac{\sum_{i=1}^{m} N_i^2 (P_i - \hat{P})}{m - 1}$$

$$q_i = 1 - p_i$$

أما حدا الثقة بمستوى ثقة $(1-\alpha)\%$:

$$\hat{P} \pm Z_{(1 - \alpha/2)} \sqrt{\hat{V}(\hat{P})} \qquad \text{.... (9 -19)}$$

تطبيق (٩ – ٧) :

باستخدام بيانات التطبيق (٩ – ٥) أوجد حدى الثقة لنسبة العمال الذين يرغبون بالانتقال إلى أماكن أخرى تعطى رواتب أعلى إذا كانت البيانات التالية للعناقيد العشرة على التوالى :

n_i = 10 , 13 , 9 , 10 , 10 , 12 , 8 , 13 , 8 , 11
N_i = 50 , 65 , 45 , 48 , 52 , 48 , 42 , 66 , 40 , 56
a_i = 4 , 5 , 2 , 3 , 5 , 3 , 3 , 4 , 2 , 4

الحـــل :

$$\hat{P} = \frac{\sum_{i=1}^{m} N_i \, p_i}{\sum_{i=1}^{m} N_i}$$

$$p_1 = \frac{4}{10} = 0.40 \,, \quad p_2 = \frac{5}{13} = 0.38 \,, \; --- \;, \; p_{10} = \frac{4}{11} = 0.36$$

إن تقدير نسبة المجتمع يساوى :

$$\hat{P} = \frac{(50 \times 0.40) + (65 \times 0.38) + ---- + (56 \times 0.36)}{50 + 65 + ---- + 56}$$

$$= \frac{176}{522} = 0.337$$

$$s_b^2 = \frac{1}{m-1} \sum_{i=1}^{m} N_i^2 \, (p_i - \hat{P})$$

$$= \frac{1}{10-1} \left[50^2 \, (0.4 - 0.337)^2 + 65^2 \, (0.38 - 0.4)^2 + ------ + 56^2 \, (0.36 - 0.337)^2 \right]$$

$$= 18.45$$

وبالتالي يكون :

$$\hat{V}(\hat{P}) = \frac{M - m}{M} \frac{(1)}{m \overline{N}^2} s_b^2 + \frac{(1}{m M \overline{N}^2} \sum_{i=1}^{m} N_i^2 \frac{(N_i - n_i)}{N_i} \frac{(p_i\ q_i)}{n_i - 1}$$

أى أن :

$$\hat{V}(\hat{P}) = \left[\frac{90 - 10}{90} \frac{1}{10 \times 50^2} \times 18.45 \right]$$

$$+ \frac{1}{10 \times 90 \times 50^2} \left[50^2 \left(\frac{(50 - 10)}{50} \frac{(0.4 \times 0.6)}{9} \right) + \right.$$

$$\left. \text{-----} + 56^2 \left(\frac{(56 - 11)}{56} \frac{(0.36 \times 064)}{10} \right) \right]$$

$= 0.00080$

ويكون حدا الثقة لنسبة الذين يرغبون فى الانتقال إلى أماكن أخرى تعطى رواتب أفضل بمستوى ثقة (٩٥٪) :

$$\hat{P} \pm Z_{(1-\alpha/2)'} \sqrt{\hat{V}(\hat{P})}$$

$$0.337 \pm 1.96 \sqrt{0.0008}$$

$$0.337 \pm 0.045$$

ويكون الحد الأدنى :

$$0.337 - 0.045 = 0.292$$

والحد الأعلى :

$$0.337 + 0.045 = 0.387$$

أى أن :

$$0.292 \leq P \leq 0.387$$

أى أن نسبة العمال الذين يرغبون فى الانتقال إلى أماكن أخرى تعطى رواتب أفضل بدرجة ثقة (٩٥٪) تتراوح بين (٢٩,٢٪) و (٣٨,٧٪) من العمال .

٩ - ٢ - ٦ تحديد حجم العينة :

يتطلب تحديد حجم العينة العنقودية ذات المرحلتين ، تحديد عدد المجموعات (العناقيد) التى يجب اختيارها (m) وذلك من بين مجموعات (عناقيد) المجتمع (M) ، ومن ثم تحديد عدد الوحدات التى يجب اختيارها من كل مجموعة ، أى تحديد (n_i) . ويعتمد تحديد (m) و (n_i) على التباين بين المجموعات ، والتباين بين المفردات داخل المجموعات ، والأساس الذى يستخدم للتحديد الأمثل لهذين الحجمين هو تحديد حجم (m) أو (n_i) حسب التباين الأكبر . فعندما تكون متوسطات المجموعات غير متجانسة بشكل كبير ، بينما تتجانس المفردات داخل المجموعات ، فإننا نختار عدداً كبيراً من المجموعات (العناقيد) أى (m) ويكون عدد المفردات المختارة من كل مجموعة (n_i) قليلاً . والعكس بالعكس عندما تكون المفردات متباينة بشكل كبير ومتوسطات المجموعات متجانسة ، نختار عدداً قليلاً من المجموعات (m) ونختار عدداً كبيراً من المفردات من كل مجموعة (n_i) .

والآن ، نرغب فى تقدير حجم (n_i . m) اللتين تجعلان تباين تقدير المتوسط $V(\hat{\bar{X}})$ أقل ما يمكن ، وذلك باستخدام تكاليف المعاينة لكل مجموعة ، وتكلفة المعاينة لكل مفردة داخل المجموعة .

لنفترض أن (c_b) هى تكلفة المعاينة لكل مجموعة (c_w) تكلفة المفردة فى المجموعة ، وأن حجم جميع العناقيد متماثلة فى الحجم ($N_1 = N_2 = --- = N_M$) ، وأن عدد الوحدات المختارة من كل عنقود متساوية أى ($n_1 = n_2 = ----- = n_m = \bar{n}$) ، نجد فى هذه الحالة أن تقدير متوسط المجتمع :

$$\hat{\bar{X}} = \frac{1}{m} \sum_{i=1}^{m} \bar{\bar{x}}_i$$

حيث :

$$\bar{\bar{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

أى متوسط العنقود (i) .

ويلاحظ أننا أهملنا الترجيح بأحجام المجموعات (العناقيد) فى المجتمع (N_i) حيث (i = 1 , 2 --- , m) وذلك لأنها متساوية .

كذلك يمكننا كتابة تباين تقدير المتوسط $V(\hat{\bar{X}})$ بافتراض أن (N_i) متساوية و (n_i) متساوية أيضاً حيث $(i = 1, 2 --- m)$:

$$V(\hat{\bar{X}}) = \frac{N - \bar{n}}{N} \frac{\sigma_{bc}^2}{m} + 1 - \frac{\bar{n}}{\bar{N}} \frac{\sigma_W^2}{\bar{n}\, m} \qquad .... (9-20)$$

حيث :

$$\sigma_{bc}^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\bar{X}_i - \bar{X})^2$$

$$\sigma_W^2 = \frac{1}{M} \sum_{i=1}^{M} \sigma_i^2$$

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2$$

وبإهمال معاملات التصحيح نجد أن :

$$V(\hat{\bar{X}}) = \frac{\sigma_{bc}^2}{m} + \frac{\sigma_W^2}{\bar{n}\; m} \qquad .... (9-21)$$

حيث : (σ_{bc}^2) هو التباين بين متوسطات المجموعات .

(σ_w^2) هو التباين بين المفردات داخل المجموعات .

ويمكننا الحصول على عدد العناقيد الأمثل (m) وحجم كل عنقود (n) باستخدام التكاليف على أساس أننا نريد تحديد الحجم الذى يعطى أكبر دقة ممكنة بأقل ما يمكن من التكاليف . إن تكاليف المعاينة تتكون من تكاليف ثابتة (c_o) وتكاليف تتوقف على عدد العناقيد الأولية (c_b) وتكاليف تتوقف على عدد الوحدات فى العنقود (c_w) أى $C = c_o + mc_b + \bar{n}\, m\, c_w$.

وسنقوم بالحصول على عدد الوحدات التى يتكون منها العنقود $(\bar{n})$ بحيث تعطى أصغر قيمة لتباين تقدير المتوسط $V(\hat{\bar{X}})$ مع ثبات التكلفة (C) أى التى تعطى أقل قيمة للتكلفة الكلية عندما يكون تباين متوسط المجتمع ثابتاً ، وللتبسيط نهمل التكاليف الثابتة أى نضع التكاليف : $C = mc_b + \bar{n}\, m\, c_w$.

والصيغة التى نستخدمها هى :

$$L = \frac{\sigma_{bc}^2}{m} + \frac{\sigma_W^2}{m\bar{n}} + \lambda \left(c_b m + c_W m \bar{n} \right)$$

وبإجراء التفاضل الجزئى لكل من $(\bar{n})$ و(m) ومساواته بالصفر نجد أن :

$$\frac{\partial L}{\partial \bar{n}} = \frac{-\sigma_W^2}{m\bar{n}^2} + \lambda c_W m = o$$

$$\frac{\partial L}{\partial m} = \frac{-\sigma_{bc}^2}{m^2} - \frac{\sigma_W^2}{\bar{n} m^2} + \lambda \left(c_b + c_W \bar{n} \right) = o$$

ومن المعادلة الأولى نجد أن :

$$\lambda = \frac{\sigma_W^2}{c_w m^2 \bar{n}^2}$$

وبالتعويض فى التفاضل الجزئى بالنسبة لـ (m) وإجراء بعض العمليات الرياضية نجد أن :

$$\frac{\sigma_{bc}^2}{m^2} + \frac{\sigma_W^2}{\bar{n} m^2} = \frac{\sigma_W^2}{c_W m^2 \bar{n}^2} \left(c_b + c_W \bar{n} \right)$$

أى أن :

$$\bar{n}^2 \sigma_{bc}^2 c_W = c_b \sigma_W^2$$

وبالتالى ·

$$\bar{n}^2 = \frac{c_b \sigma_W^2}{c_W \sigma_{bc}^2}$$

أى أن متوسط حجم العنقود الأمثل $(\bar{n})$ ولنرمز له بالرمز $(\bar{n}_{opt})$ يساوى ·

$$\bar{n}_{opt} = \sqrt{\frac{c_b}{c_W} \frac{\sigma_W^2}{\sigma_{bc}^2}} \qquad \text{.... (9 -22)}$$

ويكون إجمالي حجم العينة العنقودية ذات المرحلتين $n = \bar{n}\, m$ أي متوسط حجم العنقود في عدد العناقيد .

ويلاحظ من الصيغة (22 - 9) أن ($\bar{n}$) تتناسب طردياً مع (σ_w^2) وعكسياً مع (σ_b^2) أي أن عدد المفردات داخل كل مجموعة (عنقود) سيكون كبيراً عندما يزداد الاختلاف بين المفردات إذا قورن بالاختلاف بين المجموعات والعكس بالعكس .

ويمكننا تقدير (σ_w^2) و (σ_{bc}^2) من بيانات العينة حيث نجد أن :

$$s_W^2 = \frac{1}{m} \sum_{i=1}^{m} s_i^2$$

حيث (s_i^2) هو مقدر التباين داخل المجموعة (i) وهذا المقدر غير متحيز لـ s_w^2 . أما σ_{bc}^2 فيمكن تقديره باستخدام :

$$\hat{\sigma}_{bc}^2 = s_{bc}^2 = s^2 - \frac{s_W^2}{\bar{n}}$$

ويمكن الحصول على (s_w^2) و (s^2) من بيانات عينة استطلاعية حيث نستخدم الصيغة (22 - 9) لتحديد حجم ($\bar{n}$) وذلك باستخدام تقديرات σ_w^2 و σ_{bc}^2 كما وضحنا سابقا .

ولتقدير حجم (m) أي عدد المجموعات ، يمكننا استخدام الصيغة (21 - 9) لإيجاد الحجم الأمثل لـ m أي نستخدم :

$$V(\hat{\bar{X}}) = \frac{\sigma_{bc}^2}{m} + \frac{\sigma_W^2}{m\,\bar{n}}$$

بعد تبديل التباينات بتقديراتها الموضحة فيما سبق و ($\bar{n}$) و $(\hat{\bar{X}})$ تم تحديدها ويكون المجهول فقط هو عدد المجموعات أي العناقيد (m) .

كذلك يمكن استخدام دالة التكاليف :

$$C = mc_b + \bar{n}\, m\, c_w$$

ويكون حجم العناقيد الأمثل مساوياً في الحالتين :

$$m = \frac{\sigma_{bc}^2 + \dfrac{\sigma_W^2}{\bar{n}}}{V(\hat{\bar{X}})} \qquad \text{.... (9 -23)}$$

ونستخدم تقديرات التباين من العينة الاستطلاعية فى حال عدم توافر تباينات المجتمع ويمكن استخدام الصيغة التالية باستخدام التكاليف :

$$m = \frac{C}{c_b + \bar{n}\ c_w} \qquad .... (9 -24)$$

تطبيق (٩ – ٨) :

ترغب إحدى الوزارات فى اختيار عدد من الموظفين لتقدير متوسط درجات التقويم التى حصلوا عليها . وقد تم اختيار عينة استطلاعية من (٣) إدارات من الإدارات التى تتكون منها البالغ عددها (٢٠) إدارة وقد تم اختيار (٤) موظفين من كل من هذه الإدارات . نورد فيما يلى البيانات التى تم الحصول عليها من هذه العينة :

$$C = 900\ ,\ c_b = 200\ ,\ c_W = 4\ ,\ s_1^2 = 1.5\ ,\ s_2^2 = 2\ ,\ s_3^2 = 1\ ,\ s^2 = 3\ ,\ \bar{n} = 4$$

الحـــــل :

$$s_w^2 = \frac{1}{m} \sum_{i=1}^{m} s_i^2$$

$$= \frac{1}{3}\,(1.5 + 2 + 1) = \frac{4.5}{3} = 1.5$$

$$s_{bc}^2 = s^2 - \frac{s_W^2}{\bar{n}}$$

$$= 3 - \frac{1.5}{4} = 2.625$$

ويكون تقدير تباين المتوسط :

$$\hat{V}(\hat{\bar{X}}) = \frac{\hat{\sigma}_{bc}^2}{m} + \frac{\hat{\sigma}_W^2}{m\bar{n}}$$

$$= \frac{2.625}{3} + \frac{1.5}{3 \times 4}$$

$$= 1$$

ويكون متوسط حجم العنقود الأمثل :

$$\overline{n}_{opt} = \sqrt{\frac{c_b}{c_w} \frac{\sigma_w^2}{\sigma_{bc}^2}}$$

ونستخدم تقديرات التباين (s_w^2 , s_{bc}^2) عوضاً عن $(\sigma_w^2 + \sigma_{bc}^2)$:

$$\overline{n}_{opt} = \sqrt{\frac{200}{4} \frac{1.5}{2.625}}$$

$$= \sqrt{28.57} = 5.35$$

$$\approx 5$$

أى أن عدد الموظفين الذين سيتم اختيارهم من كل إدارة هو (٥) موظفين . ويكون عدد الإدارات المختارة أى عدد العناقيد المختارة (m) باستخدام الصيغتين (23 - 9) أو (24 - 9) :

أ – باستخدام الصيغة (23 - 9) حيث $(\overline{n} = 5)$:

$$m = \frac{2.625 + \frac{1.5}{5}}{1} = 2.925$$

$$\approx 3$$

أى ثلاث إدارات :

وباستخدام دالة التكاليف نجد أن عدد الإدارات التى سيتم اختيارها (الصيغة 24 - 9) .

$$m = \frac{900}{200 + (5 \times 1.5)} = 4.33 \approx 4$$

أى أربع إدارات .

٩ - ٢ المعاينة العنقودية ذات المراحل المتعددة :

٩ - ٢ - ١ طريقة اختيار العينة العنقودية ذات المراحل المتعددة :

كثيراً ما نحتاج إلى اختيار الوحدات النهائية فى المرحلة الثالثة أو فى مراحل أكثر من الثالثة ، وتسمى المعاينة فى هذه الحالة بالمعاينة العنقودية ذات المراحل المتعددة . ولتوضيح كيفية اختيار وحدات العينة العنقودية ذات المراحل المتعددة نورد المثال التالى :

لنفرض أن وزارة الزراعة ، ترغب فى تقدير الإنتاج الزراعى فى إحدى مناطق المملكة وأن عدد القرى فى هذه المنطقة (٥٠٠) قرية ، ولنختر كمرحلة أولى من هذه الوحدات الأولية عشوائياً (١٠) قرى . إن كل قرية تتكون من عدد من الحقول ، فيتم اختيار عدد من الحقول من كل قرية (كمرحلة ثانية) . إن كل حقل يتكون من عدد من الأقسام (المزارع) فيتم اختيار عدد من المزارع (كمرحلة ثالثة) ويتم تقدير إنتاج المزارع المختارة عن طريق حصرها حصراً شاملاً .

ويتضح من المثال السابق ، أن العينة التى سحبناها هى عينة ذات ثلاث مراحل . ويمكن أن نقسم المزرعة إلى أقسام يتم اختيار عدد منها كمرحلة رابعة وهكذا .

ويمكننا تعريف المعاينة العنقودية ذات المراحل المتعددة بأنها "عملية اختيار عينة عشوائية بسيطة من العناقيد الأولية كمرحلة أولى ومن ثم اختيار عينة عشوائية بسيطة من كل عنقود من العناقيد المختارة كمرحلة ثانية ومن ثم اختيار عينة عشوائية بسيطة من كل عنقود (من العناقيد المختارة فى المرحلة الثانية) كمرحلة ثالثة (وهكذا نتابع عملية الاختيار حسب عدد المراحل) ويتم حصر الوحدات المختارة فى المرحلة الأخيرة حصراً شاملاً وذلك للاستدلال على خصائص المجتمع .

يستخدم هذا النوع من المعاينات بشكل واسع فى التطبيقات العملية ، خاصة عندما يكون حجم المجتمع كبيراً ، ولا يتوافر إطار للوحدات النهائية . ولكن تواجهنا بعض الصعوبات عند استخدام المعاينات العنقودية ذات المراحل الثلاث أو أكثر ، وهى افتراض أن عدد وحدات العناقيد فى كل مرحلة متساو ، وهذا غير ممكن خاصة فى العناقيد الأولية ، ويستخدم فى هذه الحالة طريقة سحب الوحدات باحتمال متناسب مع الحجم .

وهناك صعوبة أخرى هى عدم التجانس بين وحدات المعاينة الأولية ، وبالمقابل وجود تجانس بين وحدات المرحلة الثانية وتجانس بين وحدات المرحلة الثالثة .

وللتخلص من هذه المشكلة ، يجب زيادة عدد الوحدات الأولية للحصول على معلومات أكثر عن وحدات المجتمع ، وهذا يؤدى إلى زيادة التكاليف خاصة إذا كانت هذه الوحدات موزعة فى منطقة واسعة . وهذا يحد من استخدام هذا النوع من العينات .

٩ - ٢ - ٢ تقديرات معالم المجتمع :

أ - تقدير متوسط المجتمع وتقدير القيمة الكلية للمجتمع :

يوجد تشابه كبير بين تقدير متوسط المجتمع والقيمة الكلية للمجتمع فى المعاينة العنقودية ذات المرحلتين والمعاينة العنقودية ذات المراحل المتعددة .

إذا استخدمنا الرموز التالية :

L عدد المجموعات (العناقيد) الأولية فى المجتمع .

L' عدد العناقيد التى تم اختيارها فى المرحلة الأولى من الوحدات الأولى .

M_i عدد الوحدات فى العنقود (i) .

m_i عدد الوحدات التى تم اختيارها من العنقود (i) الذى حجمه M_i (كمرحلة ثانية) حيث $(m_1 = m_2 = \cdots = m_L = m)$.

N_{ij} عدد الوحدات التى يتكون منها العنقود (j) الذى تم اختياره فى المرحلة الثانية .

n_{ij} عدد الوحدات التى تم اختيارها من العنقود (j) الذى يتم اختيارها كمرحلة ثالثة (نختار n_{ij} من N_{ij}) .

x_{ijk} المفردة فى الوحدة (k) من الوحدة (j) للوحدة الأولية (i) حيث لدينا (n_{ij}) مفردة .

إن الصيغة المستخدمة لتقدير القيمة الكلية للمجتمع $(\hat{X})$ هى :

$$\hat{X} = \frac{L}{L'} \sum_{i=1}^{L'} \frac{M_i}{m} \sum_{j=1}^{m} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} \qquad \text{.... (9-25)}$$

والصيغة المستخدمة لتقدير متوسط المجتمع للعينة ذات المراحل المتعددة هى :

$$\hat{\hat{X}} = \frac{1}{L'} \sum_{i=1}^{L'} \frac{M_i}{m} \sum_{j=1}^{m} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} \qquad \text{.... (9-26)}$$

حيث نعلم أن $\hat{X} = L\,\hat{\hat{X}}$ ، لذا قسمنا $\hat{X}$ على L للحصول على مقدر متوسط المجتمع الذى يعد مقدراً غير متحيز لمتوسط المجتمع .

تطبيق (٩ – ٩) :

تنتشر مخازن إحدى المؤسسات فى ثلاث مدن وترغب فى تقدير مخزونها ، فإذا كانت كل مدينة تحتوى على (٣) مخازن ، وكل مخزن يحتوى على ثلاثة أقسام . واخترنا مدينتين من المدن الثلاث ثم اخترنا مخزنين من كل منها واخترنا قسمين من كل مخزن .

المطلوب تقدير إجمالى المخزون إذا كانت مفردات المجتمع والعينة موضحة فى الجدول التالى (بمئات آلاف الريالات) :

المجتمـــع

المدينة L	المخزن M_i	القسم N_{ij}	الإنتاج X_{ijk}	القسم N_{ij}	الإنتاج X_{ijk}	القسم N_{ij}	الإنتاج X_{ijk}
1	$M_1 = 3$	$N_{11} = 3$	$X_{111} = 2$ $X_{112} = 4$ $X_{113} = 6$	$N_{12} = 3$	$X_{121} = 4$ $X_{122} = 6$ $X_{123} = 8$	$N_{13} = 3$	$X_{131} = 6$ $X_{132} = 8$ $X_{133} = 10$
2	$M_2 = 3$	$N_{21} = 3$	$X_{211} = 4$ $X_{212} = 6$ $X_{213} = 8$	$N_{22} = 3$	$X_{221} = 6$ $X_{222} = 8$ $X_{223} = 10$	$N_{23} = 3$	$X_{231} = 10$ $X_{232} = 10$ $X_{233} = 12$
3	$M_3 = 3$	$N_{31} = 3$	$X_{311} = 6$ $X_{312} = 8$ $X_{313} = 10$	$N_{32} = 3$	$X_{321} = 2$ $X_{322} = 4$ $X_{323} = 6$	$N_{33} = 3$	$X_{331} = 4$ $X_{332} = 6$ $X_{333} = 8$

العينة

$l' = 2$	$\bar{m} = 2$	$n_{ij} = 2$	x_{ijk}
$M_1 = 3$	$N_{12} = 3$ $N_{13} = 3$	$n_{11} = 2$ $n_{12} = 2$	$x_{111} = 6$, $x_{112} = 6$ $x_{121} = 4$, $x_{122} = 10$
$M_3 = 3$	$N_{31} = 3$ $N_{33} = 3$	$n_{21} = 2$ $n_{22} = 2$	$x_{211} = 10$, $x_{212} = 4$ $x_{221} = 8$, $x_{223} = 8$

الحــــــل :

من بيانات التطبيق نجد أن :

$L = 3 , L' = 2 , M_1 = M_2 = M_3 = 3 , N_{ij} = 3 , n_{ij} = 2$

– نعلم بشكل عام أن مقدر مجموع المخازن يساوى $\hat{X} = N \bar{x}$ حيث $\bar{x}$ متوسط العينة و(N) حجم المجتمع .

– لقد سحبنا عينة جزئية حجمها $n_{ij} = n = 2$ قسما من المخزن (j) فى القرية (i) ، إن مقدر إجمالى المخزون للأقسام (k) فى المخزن (j) من المدينة (i) من بيانات العينة هو :

$$x_{ij} = \sum_{k=1}^{n_{ij}} x_{ijk}$$

ومقدر متوسط المخزون للأقسام (k) فى المخزن (j) من المدينة (i) .

$$\bar{x}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$$

وهذا مقدر غير متحيز لمتوسط المخزون فى كل قسم من المخزن (j) فى المدينة (i) .
أما مقدر إجمالى المخزون لـ N_{ij} قسما فى المخزن (j) من القريـة (i) فيساوى :

$$\hat{X}_{ij} = \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} = N_{ij}\ \bar{x}_{ij}$$

وحسب بيانات التطبيق نجد أن هذه التقديرات (ولنرمز لها بالرمـوز A , B , C , D) هى :

$$x_{11} = \sum_{k=1}^{n_{11}} x_{11k} = x_{111} + x_{112} = 6 + 6 = 12 = A$$

$$x_{12} = \sum_{k=1}^{n_{12}} x_{12k} = x_{121} + x_{122} = 4 + 10 = 14 = B$$

$$x_{21} = \sum_{k=1}^{n_{21}} x_{21K} = x_{211} + x_{212} = 10 + 4 = 14 = C$$

$$x_{22} = \sum_{k=1}^{n_{22}} x_{22k} = x_{221} + x_{222} = 8 + 8 = 16 = D$$

ولتقدير إجمالى المخزون للمدن من إجمالى المخزون فى المخازن ، نعلم أن الأقسام هى عبارة عن عينة سحبت من المدن ، لذا تستعمل الصيغة $\hat{X} = N\,\bar{x}$ مرة ثانية ونجد أن :

$$\hat{X}_{ij} = N_{ij}\,\bar{x}_{ij} = \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$$

إن المتوسطات $\bar{x}_{ij}$ تساوى :

$$\bar{x}_{11} = \frac{A}{\bar{n}} = \frac{12}{2} = 6\ ,\ \bar{x}_{21} = \frac{C}{\bar{n}} = \frac{14}{2} = 7$$

$$\bar{x}_{12} = \frac{B}{\bar{n}} = \frac{14}{2} = 7\ ,\ \bar{x}_{22} = \frac{D}{\bar{n}} = \frac{16}{2} = 8$$

وبالتالى تكون قيم $(\hat{X}_{ij})$:

$$\hat{X}_{11} = \frac{3}{2} \times 12 = 18\ ,\ \hat{X}_{21} = \frac{3}{2} \times 14 = 21$$

$$\hat{X}_{12} = \frac{3}{2} \times 14 = 21\ ,\ \hat{X}_{22} = \frac{3}{2} \times 16 = 24$$

ونريد استخراج قيمة $(\hat{X}_i)$ فنجد أن :

$$\hat{\hat{X}}_i = \frac{1}{\bar{m}} \sum_{j=1}^{\bar{m}} \frac{N_{ij}}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$$

$$\hat{\hat{X}}_1 = \frac{1}{2}\left[\left(\frac{3}{2} \times 12\right) + \left(\frac{3}{2} \times 14\right)\right] = \frac{39}{2} = 19.5$$

$$\hat{\hat{X}}_2 = \frac{1}{2}\left[\left(\frac{3}{2} \times 14\right) + \left(\frac{3}{2} \times 16\right)\right] = \frac{45}{2} = 22.5$$

ب - تباين تقدير القيمة الكلية للمجتمع وتقديره :

ذكرنا فيما سبق أن تباين تقدير القيمة الكلية للمجتمع $V(\hat{X})$ للعينة العنقودية ذات المرحلتين يساوى حاصل جمع التباينين التاليين :

- التباين بين الوحدات الأولية .
- التباين بين الوحدات الثانوية (النهائية) .

ولحساب تباين تقدير القيمة الكلية للعينة ذات المراحل المتعددة ، لابد من إضافة تباين وحدات المرحلة الثالثة (أو المراحل الأخرى) ويتم ذلك على مرحلتين :

١ - جمع تباينات الوحدات الأولية مع الوحدات الثانية (وحدات المرحلة الثانية) .

٢ - جمع تباينات الوحدات الثانية مع وحدات المرحلة النهائية (الوحدات النهائية) .

ولاستخراج تباين الوحدات الأولية ووحدات المرحلة الثانية (الوحدات الثانوية) لمعاينة من ثلاث مراحل ، ولنرمز له بالرمز $V_1(\hat{X})$ ، نستخدم الصيغة التالية :

$$V_1(\hat{X}) = \left[L^2 \frac{L - L'}{L} \frac{S_b^2}{L'}\right] + \left[\frac{L}{L'} \sum_{i=1}^{L} M_i^2 \frac{M_i - \overline{m}}{M_i} \frac{S_i^2}{\overline{m}}\right] \quad .... (9\text{-}26)$$

حيث :

$$S_b^2 = \frac{1}{L-1} \sum_{i=1}^{L} (X_i - \overline{X})^2 \quad .... (9\text{-}27)$$

$$S_i^2 = \frac{1}{M_i - 1} \sum_{i=1}^{M_i} (X_{ij} - \overline{\overline{X}}_i)^2 \quad .... (9\text{-}28)$$

أما التباين فى المرحلة الثانية أى تباين وحدات المرحلة الثانية مع وحدات المرحلة الثالثة ولنرمز له بالرمز $V_2(\hat{X})$ فيساوى :

$$V_2(\hat{X}) = \left[\frac{L}{L'} \sum_{i=1}^{L} M_i^2 \frac{M_i - m}{M_i} \frac{S_i^2}{m}\right] + \left[\sum_{i=1}^{L} \frac{M_i}{m} \sum_{j=1}^{M_i} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{S_{ij}^2}{n_{ij}}\right] \quad .... (9\text{-}29)$$

حيث :

$$S_{ij}^2 = \frac{1}{N_{ij} - 1} \sum_{K=1}^{N_{ij}} (X_{ijK} - \bar{\bar{\bar{X}}}_{ij})^2 \qquad \text{.... (9 -30)}$$

$$\bar{\bar{\bar{X}}}_{ij} = \frac{1}{N_{ij}} \sum_{K=1}^{N_{ij}} X_{ijK}$$

ويكون تباين $(\hat{X})$ مساوياً لحاصل جمع التباينات فى المراحل الثلاث أى أن :

$$V(\hat{X}) = L^2 \frac{L-\ell}{L} \frac{S_1^2}{\ell} + \frac{L}{\ell} \sum_{i=1}^{L} M_i^2 \frac{M_i - \bar{m}}{M_i} \frac{S_i^2}{\bar{\bar{m}}} + \frac{L}{\ell} \sum_{i=1}^{L} \frac{M_i}{\bar{\bar{m}}} \sum_{j=1}^{M_i} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{S_{ij}^2}{n_{ij}} \qquad \text{.... (9 -31)}$$

أما مقدر $V(\hat{X})$ ولنرمز له بالرمز $\hat{V}(\hat{X})$ ، فيمكن الحصول عليه باتباع الطريقة السابقة نفسها عند استخراج $V(\hat{X})$ فى حالة المعاينة ذات المرحلتين مع إضافة تباين المرحلة الثالثة .
ونجد أن هذا التباين يساوى :

$$\hat{V}(\hat{X}) = L^2 \frac{L-\ell}{L} \frac{s_b^2}{\ell} + \frac{L}{\ell} \sum_{i=1}^{\ell} \bar{m}^2 \frac{M_i - \bar{m}}{M_i} \frac{s_i^2}{\bar{\bar{m}}} + \frac{L}{\ell} \sum_{i=1}^{\ell} \frac{M_i}{\bar{\bar{m}}} \sum_{j=1}^{\bar{m}} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{s_{ij}^2}{n_{ij}} \qquad \text{.... (9 -32)}$$

حيث :

$$s_b^2 = \frac{1}{\ell - 1} \sum_{i=1}^{\ell} (\hat{X}_i - \hat{\bar{X}})^2 \qquad \text{.... (9 -33)}$$

$$s_i^2 = \frac{1}{\bar{m} - 1} \sum_{j=1}^{\bar{m}} (\hat{X}_{ij} - \hat{\bar{X}}_i)^2 \qquad \text{.... (9 -34)}$$

$$s_{ij}^2 = \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{\bar{\bar{x}}}_{ij})^2 \qquad \text{.... (9 -35)}$$

$$\bar{\bar{\bar{x}}}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} \qquad .... (9-36)$$

ووضعنا (≡) فوق x للإشارة إلى أن العينة هى ذات ثلاث مراحل .

تطبيق (٩ – ١٠) :

باستخدام بيانات التطبيق (٩ – ٩) ، استخرج تباين تقدير القيمة الكلية للمجتمع $V(\hat{X})$.

الحـــل :

– نوجد إجمالى المخزون فى المخزن (j) من المدينة (i) أى نوجد (X_{ij}) . إن المخزون فى المخزن (j) من المدينة (i) يساوى :

$$X_{ij} = \sum_{k=1}^{N_{ij}} X_{ijK}$$

$$X_{11} = \sum_{k=1}^{N_{11}} X_{11K} = X_{111} + X_{112} + X_{113}$$

$$= 2 + 4 + 6 + = 12$$

وبالطريقة نفسها نستخرج X_{ij} فنجد أن :

$X_{11} = 12$, $X_{12} = 18$, $X_{13} = 24$
$X_{21} = 18$, $X_{22} = 24$, $X_{23} = 30$
$X_{31} = 24$, $X_{32} = 12$, $X_{33} = 18$

– نوجد X_i أى إجمالى المخزون فى المدينة (i) . إن :

$$X_i = \sum_{j=1}^{M_i} X_{ij}$$

وبالتالى نجد أن :

$$X_1 = \sum_{j=1}^{M_1} X_{1j} = X_{11} + X_{12} + X_{13}$$

$$= 12 + 18 + 24 = 54$$

وباستخدام الطريقة نفسها نستخرج X_i حيث نجد أن :

$$X_1 = 54 \ , \ X_2 = 72 \ , \ X_3 = 54$$

- إيجاد متوسط المخزون فى المدينة $(\overline{X})$:

$$\overline{X} = \frac{1}{L}\sum_{i=1}^{L} X_i$$

$$= \frac{1}{3}(54 + 72 + 54) = 60$$

- إيجاد قيمة (S_b^2) :

$$S_b^2 = \frac{1}{L-1}\sum_{i=1}^{L}(X_i - \overline{X})^2$$

$$= \frac{1}{3-1}\left[(54-60)^2 + (72-60)^2 + (54-60)^2\right]$$

$$= 108$$

- إيجاد متوسط المخزون فى كل مخزن $(\overline{\overline{X}}_i)$:

$$\overline{\overline{X}}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} X_{ij}$$

$$\overline{\overline{X}}_1 = \frac{1}{M_1}\sum_{j=1}^{M_1} X_{ij} = \frac{1}{M_1}(X_{11} + X_{12} + X_{13})$$

$$= \frac{1}{3}(12 + 18 + 24) = 18$$

$$\overline{\overline{X}}_2 = \frac{1}{3}(18 + 24 + 30) = 24$$

$$\bar{\bar{X}}_3 = \frac{1}{3}(24 + 12 + 18) = 18$$

إيجاد التباين (S_i^2) :

$$S_i^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}(X_{ij} - \bar{\bar{X}}_i)^2$$

$$S_1^2 = \frac{1}{3 - 1}\left[(12 - 18)^2 + (18 - 18)^2 + (24 - 18)^2\right] = 36$$

وبالطريقة نفسها نستخرج S_i^2 لجميع قيم (i) فنجد أن :

$S_1^2 = 36$, $S_2^2 = 36$, $S_3^2 = 36$

- إيجاد متوسط المخزون للقسم (i) في المخزن (j) $(\bar{\bar{X}}_{ij})$:

$$\bar{\bar{X}}_{ij} = \frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} X_{ijK} = \frac{X_{ij}}{N_{ij}}$$

$$\bar{\bar{X}}_1 = \frac{X_{11}}{N_{11}} = \frac{12}{3} = 4$$

ونستخرج باقي القيم بالطريقة نفسها حيث نجد أن :

$$\bar{\bar{X}}_{11} = 4 , \bar{\bar{X}}_{12} = 6 , \bar{\bar{X}}_{13} = 8$$

$$\bar{\bar{X}}_{21} = 6, \bar{\bar{X}}_{22} = 8 , \bar{\bar{X}}_{23} = 10$$

$$\bar{\bar{X}}_{31} = 8 , \bar{\bar{X}}_{32} = 4 , \bar{\bar{X}}_{33} = 6$$

- إيجاد التباين (S_{ij}^2) :

$$S_{ij}^2 = \frac{1}{N_{ij} - 1}\sum_{k=1}^{N_{ij}}(X_{ijk} - \bar{\bar{X}})^2$$

$$S_{11}^2 = \frac{1}{N_{11} - 1}\sum_{k=1}^{N_{11}}(X_{11k} - \bar{\bar{X}}_{11})^2$$

$$S_{11}^2 = \frac{1}{3-1}\left[(2-4)^2 + (4-4)^2 + (6-4)^2\right] = 4$$

وبالتالى نجد أن :

$$S_{11}^2 = 4 , S_{12}^2 = 4 , S_{13}^2 = 4 , S_{21}^2 = 4$$

$$S_{22}^2 = 4 , S_{23}^2 = 4 , S_{31}^2 = 4 , S_{32}^2 = 4 , S_{33}^2 = 4$$

– استخراج تباين تقدير القيمة الكلية $V(\hat{X})$:

الحد الأول من الصيغة (3 - 9) يساوى :

$$L^2 \frac{L-\ell}{L} \frac{S_b^2}{\ell} = 3^2 \times \frac{3-2}{3} \times \frac{108}{2} = 162$$

– الحد الثانى يساوى :

$$\frac{L}{\ell} \sum_{i=1}^{L} M_i^2 \frac{M_i - \overline{m}}{M_i} \frac{S_i^2}{\overline{\overline{m}}} = \frac{3}{2} \left\{\left[3^2 \times \frac{3-2}{3} \times \frac{36}{2}\right] + \left[3^2 \times \frac{3-2}{3} \times \frac{36}{2}\right] + \right.$$

$$\left. \left[3^2 \times \frac{3-2}{3} \times \frac{36}{2}\right]\right\} = 243$$

– الحد الثالث يساوى :

$$\frac{L}{\ell} \sum_{i=1}^{L} \frac{M_i}{\overline{\overline{m}}} \sum_{j=1}^{M_i} N_{ij}^2 \frac{N_{ij} - n_{ij}}{N_{ij}} \frac{S_{ij}^2}{n_{ij}}$$

$$= \frac{L}{\ell}\left[\frac{M_1}{\overline{\overline{m}}}\left(N_{11}^2 \frac{N_{11} - n_{11}}{N_{11}} \frac{S_{11}^2}{n_{11}} + N_{12}^2 \frac{N_{12} - n_{12}}{N_{12}} \frac{S_{12}^2}{n_{12}}\right.\right.$$

$$\left. + N_{13}^2 \frac{N_{13} - n_{13}}{N_{13}} \frac{S_{13}^2}{n_{13}}\right) + \frac{M_2}{\overline{\overline{m}}}\left(N_{21}^2 \frac{N_{21} - n_{21}}{N_{21}} \frac{S_{21}^2}{n_{21}}\right.$$

$$+ \cdots + \cdots \Big) + \frac{M_3}{\overline{m}} \; N_{31}^2 \; \frac{N_{31} - n_{31}}{N_{31}} \; \frac{S_{31}^2}{n_{31}} + \cdots + \cdots \Big)\Big]$$

$$= \frac{3}{2}\left[\frac{3}{2}\left(9 \times \frac{3-2}{3} \times \frac{4}{2}\right) \times 9\right] = 81$$

وضربت بـ (٩) لأنها مكررة (٩) مرات .

ويكون تباين تقدير القيمة الكلية مساوياً لمجموع الحدود الثلاثة :

$$V(\hat{X}) = 162 + 243 + 81 = 486$$

تطبيق (٩ – ١١) :

باستخدام بيانات ونتائج التطبيق (٩ – ٩) ، أوجد تقدير تباين تقدير القيمة الكلية للمجتمع $\hat{V}(\hat{X})$.

الحـــــل :

لدينا البيانات والنتائج التالية :

$$L = 3 \,,\, \acute{L} = 2 \,,\, M_1 = M_2 = M_3 = 3 \,,\, N_{ij} = 3 \,,\, n_{ij} = 2 \,,\, \overline{m} = 2$$

$$x_{11} = 12 \,,\; x_{12} = 14 \,,\; x_{21} = 14 \,,\; x_{22} = 16$$

$$\bar{x}_{11} = 6 \,,\; \bar{x}_{12} = 7 \,,\; \bar{x}_{21} = 7 \,,\; \bar{x}_{22} = 8$$

$$\hat{X}_{11} = 18 \,,\; \hat{X}_{12} = 21 \,,\; \hat{X}_{21} = 21 \,,\; \hat{X}_{22} = 24$$

$$\hat{\hat{X}}_1 = 19.5 \,,\; \hat{\hat{X}}_2 = 22.5 \,,\; \hat{\hat{X}} = 63 \,,\; \hat{X}_1 = 58.5$$

$$\hat{X}_2 = 67.5 \,,\; \hat{X} = 189$$

– إيجاد التباين (s_b^2) :

$$s_b^2 = \frac{1}{\acute{L}-1} \sum_{i=1}^{\acute{L}} (\hat{X}_i - \hat{\hat{X}})^2$$

$$= \frac{1}{2-1}\left[(58.5 - 63)^2 + (67.5 - 63)^2\right] = 40.5$$

- إيجاد التباين (s_i^2) :

$$s_i^2 = \frac{1}{\overline{m} - 1} \sum_{j=1}^{\overline{m}} (\widehat{X}_{ij} - \widehat{\overline{X}}_i)^2$$

$$s_1^2 = \frac{1}{2 - 1} \left[(18 - 19.5)^2 + (21 - 19.5)^2 \right] = 4.5$$

$$s_2^2 = \frac{1}{2 - 1} \left[(21 - 22.5)^2 + (24 - 22.5)^2 \right] = 4.5$$

- إيجاد التباين (s_{ij}^2) :

$$s_{ij}^2 = \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (x_{ijk} - \overline{\overline{\overline{x}}}_{ij})^2$$

حيث : $\overline{\overline{\overline{x}}}_{ij} = \overline{x}_{ij}$ فى التطبيق :

$$s_{11}^2 = \frac{1}{2 - 1} \left[(6 - 6)^2 + (6 - 6)^2 \right] = 0$$

$$s_{12}^2 = \frac{1}{2 - 1} \left[(4 - 7)^2 + (10 - 7)^2 \right] = 18$$

$$s_{21}^2 = \frac{1}{2 - 1} \left[(10 - 7)^2 + (4 - 7)^2 \right] = 18$$

$$s_{22}^2 = \frac{1}{2 - 1} \left[(8 - 8)^2 + (8 - 8)^2 \right] = 0$$

- لنستخرج تقدير التباين $\hat{V}(\hat{X})$:

- الحد الأول يساوى :

$$L^2 \frac{L - l'}{L} \frac{S_b^2}{l'} = 3^2 \times \frac{3 - 2}{3} \times 40.5 = 121.5$$

- الحد الثانى يساوى

$$\frac{L}{l'} \sum_{i=1}^{l'} \overline{m}^2 \frac{M_i - \overline{m}}{M_i} \frac{s_i^2}{\overline{m}} = \frac{3}{2} \left[(2^2 \times \frac{3-2}{3} \times \frac{4.5}{2}) + (2^2 \times \frac{3-2}{3} \times \frac{4.5}{2}) \right] = \frac{3}{2} \left[3 + 3 \right] = 9$$

- الحد الثالث يساوى :

$$\frac{L}{\ell}\sum_{i=1}^{\ell}\frac{M_i}{\overline{m}}\sum_{j=1}^{\overline{m}}N_{ij}^2\ \frac{N_{ij}-n_{ij}}{n_{ij}}\ \frac{s_{ij}^2}{n_{ij}}$$

$$=\frac{3}{2}\left[\ \frac{3}{2}\ (3^2\ x\ \frac{3-2}{3}\ x\ \frac{0}{2})+(3^2\ \ \frac{3-2}{3}\ x\ \frac{18}{2}\ )\right.$$

$$\left.+\frac{3}{2}\ (3^2\ x\ \frac{3-2}{3}\ x\ \frac{18}{2})+(3^2\ \frac{3-2}{3}\ x\ \frac{0}{2}\ )\ \right]$$

$$=\frac{3}{2}\left[\ \frac{3}{2}\ (0+27)+\frac{3}{2}\ (27+0)\ \right]-40.5$$

ويكون تقدير تباين تقدير القيمة الكلية للمجتمع $\hat{V}(\hat{X})$ مساوياً :

$$\hat{V}\ (\hat{X}\ )=121.5+9+40.5=171$$

وهو المطلوب .

٩ - ٤ المعاينة الطبقية العنقودية : (Stratified Cluster Sampling)

ذكرنا فيما سبق عند دراستنا للمعاينة الطبقية ، أننا قسمنا المجتمع إلى طبقات تحتوى كل منها على وحدات متجانسة إلى حد ما فيما بينها ، بينما تختلف الطبقات من واحدة لأخرى . ويوجد نوع آخر من المعاينات يجمع بين المعاينة الطبقية والمعاينة العنقودية ويسمى "المعاينة الطبقية العنقودية" ، حيث يقسم المجتمع إلى طبقات وتعد كل طبقة من هذه الطبقات كمجتمع صغير يتألف من عدد من العناقيد . ونطبق طريقة اختيار المعاينة العنقودية على كل طبقة ومن ثم نقوم باستخراج التقديرات المطلوبة .

ويستخدم هذا النوع من المعاينات بشكل واسع ، مثلاً لمعرفة آراء السكان فى دولة معينة ، يمكننا تقسيم مناطق الدولة إلى (L) مدينة وقرية (طبقة) ونقسم كل طبقة إلى أحياء (عناقيد) عددها $(\overline{M})$ عنقوداً يمثل كل حى منها منطقة انتخابية . وتتم عملية المعاينة بسحب عنقود حى أو أكثر $(\overline{m}\geq 1)$ من كل طبقة (أى من كل مدينة أو قرية) ، ثم نختار عدداً من الأشخاص من كل حى من الأحياء المختارة $(\overline{n})$.

إن تقدير مجموع المجتمع $(\hat{X})$ يساوى العلاقة (24 - 9) مع إدخال بعض التعديلات حيث إن (L) ترمز إلى عدد العناقيد الأولية . وإذا افترضنا أن (L) ترمز إلى عدد الطبقات و(L') ترمز إلى عدد العناقيد ، فإننا نحصل على تقدير القيمة الكلية للمجتمع للمعاينة الطبقية العنقودية ويساوى :

$$\hat{X} = \sum_{h=1}^{L} \frac{M_h}{m_h} \sum_{i=1}^{m_h} \frac{N_{hi}}{n_{hi}} \sum_{j=1}^{n_{hi}} x_{hij} \quad \text{.... (9 -37)}$$

حيث $(\hat{X})$ هو تقدير غير متحيز للقيمة الكلية للمجتمع .

وبإجراء التعديلات نفسها على تباين العينة العنقودية نجد أن تباين العينة الطبقية العنقودية يساوى :

$$V(\hat{X}) = \sum_{h=1}^{L} M_h^2 \frac{M_h - m_h}{M_h} \frac{S_h^2}{m_h} + \sum_{h=1}^{L} \sum_{i=1}^{m_h} N_{hi}^2 \frac{N_{hi} - n_{hi}}{N_{hi}} \frac{S_{hi}^2}{n_{hi}} \quad \text{. (9 -38)}$$

حيث :

$$S_h^2 = \frac{1}{M_h - 1} \sum_{i=1}^{M_h} (X_{hi} - \overline{X}_h)^2$$

$$S_{hi}^2 = \frac{1}{N_{hi} - 1} \sum_{j=1}^{N_{hi}} (X_{hij} - \overline{\overline{X}}_{hi})^2$$

ويمثل (S_h^2) تباين العناقيد الأولية فى الطبقة (h) بينما يمثل (S_{hi}^2) تباين الوحدات بين الوحدات للعنقود (i) فى الطبقة (h) . ويلاحظ أن التباين بين العناقيد (الطبقات فى هذه الحالة) قد حذف من الصيغة (32 - 9) .

٩ – ٥ المعاينة العنقودية باحتمالات متناسبة مع الحجم :

إذا كانت أحجام جميع العناقيد (N_i) معلومة ، نستطيع اختيار العناقيد بحيث يكون لكل عنقود الفرصة فى الظهور باحتمال يتناسب مع حجمه ويؤدى ذلك إلى زيادة الدقة فى التقدير .

إذا رمزنا إلى احتمال اختيار العنقود (i) فى العينة العنقودية بالرمز π_i حيث ($\pi_i = N_i/N$) فإن مقدر مجموع قيم المجتمع :

$$\hat{T} = \frac{N}{n} \sum_{i=1}^{n} \bar{x}_i \qquad \text{.... (9 - 39)}$$

ويكون مقدر متوسط المجتمع .

$$\hat{\mu} = \frac{N}{m} \sum_{i=1}^{m} \bar{x}_i \qquad \text{.... (9 - 40)}$$

حيث ($\bar{x}_i$) هو متوسط قيم المفردات فى العنقود (i) فى العينة ، ويكون تباين تقدير متوسط المجتمع $V(\hat{\mu})$ وتقديره $\hat{V}(\hat{\mu})$ وتباين تقدير القيمة الكلية $V(\hat{T})$ وتقديره $\hat{V}(\hat{T})$ كما يلى

$$V(\hat{\mu}) = \frac{1}{mN} \sum_{i=1}^{M} N_i (\mu_i - \mu)^2$$

$$\hat{V}(\hat{\mu}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{x}_i - \hat{\mu})^2$$

$$V(\hat{T}) = \frac{1}{mN} \sum_{i=1}^{M} N_h (\mu_i - \mu)^2 \qquad \text{.... (9 - 41)}$$

$$\hat{V}(\hat{T}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{x}_i - \hat{\mu})^2$$

ويمكننا استخراج حدود الثقة بمستوى ثقة %(α - 1) باستخدام الطرق السابقة نفسها .

تطبيق (٩ – ١٢) :

يرغب معهد الإدارة العامة فى تقدير متوسط عدد أيام غياب المتدربين فى (٥) برامج تدريبية فإذا كانت لدينا البيانات التالية :

رقم البرنامج	١	٢	٣	٤	٥	المجموع
	٢٠	٢٥	١٥	٢٠	٢٥	١٠٥

وضح كيفية اختيار عينة عنقودية متناسبة مع الحجم مكونة من (٣) برامج . ثم أوجد تقدير متوسط عدد أيام الغياب بدرجة ثقة ٩٥٪ .

الحـــل :

نوجد المجموع المتجمع لعدد الموظفين والمدى المتجمع .

رقم البرنامج	المجموع التراكمي	المدى المتجمع
١	٢٠	١ – ٢٠
٢	٤٥	٢١ – ٤٥
٣	٦٠	٤٦ – ٦٠
٤	٨٠	٦١ – ٨٠
٥	١٠٥	٨١ – ١٠٥

ونريد اختيار ثلاثة أرقام عشوائية (باستخدام جداول الأرقام العشوائية) تقع بين (١) و (١٠٥) ، ولنفترض أن الأرقام المختارة هى (٧٥) ، (٤٠) ، (٩٠) أى أننا اخترنا البرامج ذوات الأرقام ٤ ، ٢ ، ٥ ونقوم بتدوين عدد أيام الغياب (من سجلات إدارة التسجيل) فتكون على التوالى :

رقم البرنامج	٤	٢	٥	المجموع
عدد أيام الغياب (X_i)	١٠٠	٧٥	١٠٠	٢٧٥
عدد المتدربين (N_i)	٢٠	٢٥	٢٥	٧٠

ويكون تقدير متوسط غياب كل فصل :

$$\bar{x}_1 = \frac{100}{20} = 5 , \quad \bar{x}_2 = \frac{75}{25} = 3 , \quad \bar{x}_3 = \frac{100}{25} = 4$$

أما تقدير متوسط أيام الغياب للمتدرب فيساوى :

$$\hat{\mu} = \frac{1}{3} \quad (5 + 3 + 4) = 4$$

أما تقدير تباين هذا التقدير فيساوى :

$$\hat{V}(\hat{\mu}) = \frac{1}{3(3-1)} \left[(5-4)^2 + (3-4)^2 + (4-4)^2)\right]$$

$$= \frac{1}{6} [1 + 1 + 0) = 0.333$$

وتكون حدود الثقة بمستوى ثقة ٩٥٪ :

$$4 \pm 1.96 \sqrt{0.333}$$

أى أن

$$2.87 \leq \mu \leq 5.13$$

أى أن متوسط أيام الغياب للمتدرب فى جميع البرامج التدريبية يتراوح بين (٢,٨٧) من الأيام و(٥,١٣) منها بدرجة ثقة (٩٥٪) .

الفصل
العاشر

أنواع المعاينات الأخرى

استعرضنا فيما سبق أهم المعاينات الاحتمالية وغير الاحتمالية ، وهناك بعض الأنواع الأخرى للمعاينات التي تستخدم أحياناً في المجالات العملية ، وتعتمد طرق معالجة بعض المعاينات على الطرق التي استخدمت في المعاينة التي تمت دراستها ، ويتطلب بعضها الآخر معالجتها بطرق أخرى وأهم هذه المعاينات :

- المعاينة المزدوجة .
- المعاينة في مناسبتين أو أكثر .
- المعاينة المساحية .
- المعاينات للمجتمعات البرية .

وسنقوم باستعراض هذه الأنواع باختصار .

١٠ - ١ المعاينة المزدوجة (Double Sampling)

يفضل بعض الباحثين في بعض الحالات ، جمع بيانات معينة عن بعض الوحدات الإحصائية المختارة بأسلوب المعاينة ثم اختيار عينة جزئية من العينة الأصلية لدراسة الخاصية قيد الدراسة ، وتستعمل العينة الجزئية لإيجاد التقديرات الإحصائية . وتسمى المعاينة التي تسحب عن طريق أخذ عينة كبيرة للحصول على معلومات إضافية بتكاليف قليلة ثم اختيار عينة صغيرة من العينة الكبيرة لدراسة الظاهرة المطلوبة بالمعاينة المزدوجة .

وتستخدم بيانات العينة الكبيرة لتقدير معالم ظاهرة ما ، (ولنرمز للمتغيربالرمز X) خاصة وسطها الحسابي باستخدام عدة طرق للتقدير :

- التقدير بالانحدار .
- التقدير بالنسبة .
- التقدير بتقسيم المجتمع إلى طبقات .

أما البيانات التي نحصل عليها من العينة الفرعية الصغيرة والتي تكون تكاليفها قليلة ، فتستخدم مع البيانات التي جمعت في العينة الكبيرة وتقديراتها لتقدير معالم الظاهرة المدروسة (Y) .

ولتوضيح هذا النوع من المعاينات ، نفترض أن لدينا مجتمعاً يتكون من (N) موظفاً يعملون في إدارة الرقابة المالية ، ونرغب في تقدير متوسط عدد أوامر الصرف التي يدققونها ومتوسط مبالغها وأنواعها ... إن اختيار عينة لجمع البيانات المطلوبة من الموظفين يتطلب تكاليف ضخمة ، لذا يمكننا اختيار عينة كبيرة الحجم من الموظفين بجمع بيانات عن عدد

أوامر الصرف ومبالغها التي تم تدقيقها في العام الماضي ، وهذه البيانات متوافرة ولاتتطلب تكاليف كبيرة ، وبذلك نحصل على معلومات لها علاقة بالبيانات للسنة الحالية ، ثم نختار عينة صغيرة الحجم من وحدات العينة الكبيرة الحجم ، ونجمع بيانات عن الظاهرة ، ونقوم بإجراء التحليلات والتقديرات المطلوبة باستخدام طرق التقدير بالانحدار أو بالنسبة أو بغيرهما .

وسنقوم فيما يلي بشرح مختصر لهذا النوع من المعاينات باستخدام طرق التقدير الثلاث المشار إليها فيما سبق .

١٠-١-١ التقدير بالانحدار في المعاينة المزدوجة

(Regression Estimate in Double Sampling)

تعتمد طريقة التقدير بالانحدار على استخدام المعلومات الإضافية (المتممة) عن طريق المتغير المساعد (Y) الذي يرتبط مع المتغير (X) ارتباطاً قوياً .

لنفترض أن لدينا مجتمعاً عدد مفرداته (N) ونريد تقدير متوسط المجتمع للظاهرة (X) . ونظراً لارتفاع تكاليف المعاينة ، نستخدم طريقة التقدير بالانحدار في المعاينة المزدوجة للحصول على التقدير المطلوب وفق الخطوات التالية :

- نقوم باختيار عينة كبيرة قليلة التكاليف حجمها (n') من المجتمع لقياس المتغير (Y) فنحصل على القيم ($y_1 \; y_2 \; ... \; y_{n'}$) .
- نختار عينة فرعية حجمها (n) من العينة الكبيرة (n') ونقيس الظاهرة (X) فنحصل على القيم (x_i) والتي تقابلها قيم من (y_i) عن كل مفردة من العينة الفرعية أي نحصل على $(x_1 , y_1) , (x_2 , y_2) \ldots\ldots (x_n , y_n)$.
- إن متوسط العينة الكبيرة ولنرمزله بالرمز ($\bar{y}'$) ومتوسط العينة الصغيرة ($\bar{y}$) يساويان :

$$\bar{y}' = \frac{1}{n'} \sum_{i=1}^{n'} y_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

فيكون مقدر متوسط المتغير (X) باستخدام خط الانحدار :

$$\hat{\mu}_x = \bar{x} + \hat{B}\left(\bar{y}' - \bar{y}\right) \qquad .... (10 - 1)$$

حيث $\bar{x}$ هو متوسط المتغير (X) من العينة الفرعية ويساوى :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

و$(\hat{B})$ معامل انحدار (X) على (Y) محسوباً من بيانات العينة الفرعية ويساوى .

$$\hat{B} = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

وهذا المقدر متحيز بمقدار التغاير $- \text{Cov}\,(\hat{B}, \bar{y})$ من الدرجة $\left(\frac{1}{n}\right)$ ولكنه متسق ، ويمكن إهمال التحيز إذا كان حجم العينة كبيراً .

أما مقدر تباين تقدير متوسط المجتمع فيساوى (تقريباً) :

$$\hat{V}\left(\hat{\mu}_x\right) = \frac{s_e^2}{n} + \frac{s_x^2 - s_e^2}{n'} - \frac{S_x^2}{N} \qquad .... (10 - 2)$$

حيث

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

$$S_e^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 - \hat{B}^2 \sum_{i=1}^{n}(y_i-\bar{y})\right]$$

ويمكن إهمال الحد الأخير فى التباين $\hat{V}(\hat{\mu}_L)$ إذا كان حجم المجتمع غير معلوم لكبر حجمه .

تطبيق (١٠ – ١)

يرغب معهد الإدارة العامة فى تقدير مستوى المتقدمين لبرنامج الحاسب الآلى فى مادة الرياضيات البالغ عددهم (١٥٠٠) طالب وقد تم اختيار (٢٠٠) طالب عشوائيًا وتبين أن متوسط درجات الطالب فى مادة الرياضيات فى الشهادة الثانوية (٧٨) درجة .

وتقرر إجراء اختبار القبول لـ(٥٠) متقدمًا فى مادة الرياضيات تم اختيارهم من بين العينة التى تم اختيارها ، وتبين لنا أن متوسط درجات الطالب فى الاختبار (٧١) درجة ومتوسط درجاتهم فى الثانوية العامة (٧٤) درجة . المطلوب تقدير متوسط درجة الطالب المستجد فى مادة الرياضيات بطريقة الانحدار علمًا بأن .

$$\sum_{i=1}^{50}(y_i-\bar{y})^2 = 400\ ,\ \sum_{i=1}^{50}\left(x_i-\bar{x}\right) = 810\ ,\ \sum_{i=1}^{50}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right) = 500$$

الحـل

لدينا البيانات التالية :

$$N = 1500\ ,\ n' = 200\ ,\ n = 50$$

$$\bar{y} = 74\ ,\ \bar{y}' = 78\quad \bar{x} = 71$$

حيث يرمز المتغير (y') إلى درجات الطلاب فى الثانوية العامة (العينة الكبيرة)

المتغير (y) إلى درجات الطلاب فى الثانوية العامة (العينة الصغيرة)

المتغير (x) إلى درجات الطلاب فى اختبار القبول (العينة الصغيرة)

ونستخرج التقديرات التالية :

$$\hat{B} = \frac{\sum_{i=1}^{50} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{50} (y_i - \bar{y})^2}$$

$$= \frac{500}{400} = 1.25$$

ويكون تقدير متوسط الدرجات

$$\hat{\mu}_x = 71 + 1.25\ (78 - 74) = 76$$

– أما تقدير إجمالى التباين لتقدير متوسط المجتمع فيساوى

$$\hat{V}\left(\hat{\mu}_x\right) = \frac{S_e^2}{n} + \frac{S_x^2 - S_e^2}{n'} - \frac{S_x^2}{N}$$

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{50} (x_i - \bar{x})^2$$

$$= \frac{1}{50-1} (810) = 16.53$$

$$S_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2 - \hat{B}^2 \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2$$

$$= \frac{1}{50-2} \left[810 - (1.25)^2 (400) \right] = \frac{185}{48} = 3.85$$

ويكون تقدير التباين :

$$\hat{V}\left(\hat{\mu}_x\right) = \frac{3.85}{50} + \frac{16.53 - 3.85}{500} - \frac{16.88}{1500}$$

$$= 0.077 + 0.063 - 0.011 = 0.130$$

ويمكننا إهمال الحد الأخير إذا كان حجم المجتمع مجهولاً لكبر حجمه . ويمكننا استخراج حدى الثقة وذلك باتباع الطرق التى تم شرحها فيما سبق .

١٠-١-٢ التقدير بالنسبة فى المعاينة المزدوجة .

(Ratio Estimate in Double Sampling)

يستخدم بعض الإحصائيين التقديرات التى تتكون من النسبة بين متغيرين لتقدير معالم المجتمع وذلك عن طريق المعاينة المزدوجة . إن الغرض من استخدام طريقة التقدير بالنسبة فى المعاينة المزدوجة هى الحصول على دقة أعلى باستخدام الارتباط بين المتغيرين (X) (Y) من بيانات العينة .

وكمثال على هذا النوع من التقديرات ، لنفترض أننا نرغب فى تقدير الدخل لموظفى إحدى الجهات . إذا اخترنا عينة من الموظفين ذات حجم كبير واستخرجنا نسبة الدخل إلى ظاهرة أخرى ذات علاقة قوية بينهما كالإيجار أو الإنفاق الشهرى على الغذاء ، فإننا نستخدم هذه

النسبة فى تقدير متوسط الدخل لعينة من الموظفين يتم اختيارهم من العينة الكبيرة ، ويتم فيها تقدير متوسط الدخل باستخدام البيانات الإضافية المتاحة من العينة الأولى . ولتوضيح كيفية التقدير بالنسبة فى المعاينة المزدوجة نتبع الخطوات التالية :

- نختار عينة عشوائية كبيرة الحجم حجمها (n') ويكون متوسطها ($\bar{y}'$) تقدير للمتوسط μ_y الخاص بالظاهرة (Y) .
- نقوم باختيار عينة فرعية حجمها (n) وحدة ، ونقيس الظاهرة (x) فنحصل على قيم المتغير (x) والقيم المقابلة لها فى العينة الكبيرة أى يكون لدينا أزواج من القيم (y_i , x_i) وبالتالى نحسب متوسطاتها $\bar{y} , \bar{x}$.
- إن تقدير النسبة للمتوسط μ_x يتم استخراجه بضرب نسبة ($\bar{x}$) إلى ($\bar{y}$) ومن ثم ضربه بمتوسط العينة الكبيرة ($\bar{y}'$) أى .

$$\hat{\mu} = \frac{\bar{x}}{\bar{y}} \quad \bar{y}' \qquad \text{.... (10 - 3)}$$

حيث تمثل $\left(\frac{\bar{x}}{\bar{y}}\right)$ النسبة بين المتوسطين ($\hat{R}$) .

أما تقدير التباين التقريبى لتقدير متوسط المجتمع فيمكن استخراجه من تقديرات العينتين باستخدام الصيغة التالية :

$$\hat{V}\left(\hat{\mu}\right) = \frac{1}{n}\left(s_x^2 - 2\hat{R}\, s_{xy} + \hat{R}^2 s_y^2\right) + \frac{1}{n'}\left(2\hat{R}\, s_{xy} - \hat{R}^2 s_y^2\right) - \frac{S_x^2}{N} \qquad \text{.... (10 - 4)}$$

ويمكن أهمال الحد الأخير إذا كان حجم المجتمع مجهولاً ولكونه كبيراً حيث (s^2_y) و(s^2_x) هما تباين المتغيرين (Y) و(X) من العينة و(s_{xy}) هو التغاير للمتغيرين و$\hat{R}$ نسبة المتوسطين من العينة .

تطبيق (١٠ – ٢)

باستخدام بيانات التطبيق (١٠–١) أوجد تقدير المتوسط باستخدام طريقة التقدير بالنسبة .

الحـل :

لدينا البيانات التالية :

$$N = 1500, n' = 200 , n = 50 , \overline{y} = 74 , \overline{y}' = 78$$

$$\overline{x} = 71 , s^2_x = 16.88$$

– إن تقدير متوسط درجات المتقدم للاختبار باستخدام طريقة التقدير بالنسبة يساوى :

$$\hat{\mu} = \frac{\overline{x}}{\overline{y}} \quad \overline{y}'$$

$$= \frac{71}{74} \text{ x } 78 = 74.84$$

ونجد أن تقدير النسبة يساوى :

$$\hat{R} = \frac{71}{74} = 0.959$$

- لاستخراج تقدير التباين التقريبي نستخدم الصيغة (4 - 10) لذا نحتاج إلى حساب :

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{50-1} (400) = 8.16$$

$$s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y})$$

$$= \frac{1}{50-1} \times 500 = 10.20$$

ويكون تقدير التباين التقريبي لتقدير متوسط المجتمع

$$\hat{V} (\hat{\mu}) = \frac{1}{50} \left[16.88 - (2 \times 0.959 \times 10.20) + (0.959 \times 8.16) \right]$$

$$+ \frac{1}{200} \left[(2 \times 0.959 \times 10.20) - (0.959 \times 8.16) \right] + \frac{16.88}{1500}$$

$$= \frac{1}{50} (16.88 - 19.56 + 7.83) + \frac{1}{200} (19.56 - 7.83) + 0.0113$$

$$= 0.103 + 0.059 + 0.0113 = 0.1733$$

ويمكن استخراج حدى الثقة باتباع الطرق السابقة الموضحة عند دراسة الأنواع الأخرى للمعاينات .

١٠-١-٢ التقدير بتقسيم المجتمع إلى طبقات فى المعاينة المزدوجة :

(Stratified Estimate in Double Sampling)

يتم فى هذه الطريقة ، تقدير متوسط المجتمع وتقدير تباينه عن طريق تقسيم المجتمع إلى طبقات فى العينتين الكبيرة والصغيرة حيث نتبع الخطوات التالية :

- يتكون المجتمع من (N) مفردة وقسمناه إلى (L) طبقة حسب المتغير (X)
- اخترنا عينة عشوائية كبيرة حجمها (n') وتكون :

نسبة مفردات المجتمع فى الطبقة (h) : $W_h = \frac{N_h}{N}$

- نسبة مفردات العينة الكبيرة فى الطبقة (h) : $W'_h = \frac{n'_h}{n}$

حيث (N_h) عدد مفردات المجتمع للطبقه (h) و(n_h) عددمفردات العينة للطبقة (h) ويمكن اعتبار (w'_h) كمقدر لنسبة مفردات المجتمع فى الطبقة (h) وهو مقدر غير متحيز لهذه النسبة .

- نختار عينة فرعية طبقية حجمها (n)مفردة منها (n_h) من الطبقة (h) مسحوبة من مفردات العينة الأولى فى الطبقة (h) أى من (n'_h) .
- نريد تقدير نسب الطبقات من العينة الأولى(w'_h) ومن ثم تقدير متوسط الطبقات ($\overline{x}_h$) وذلك لجميع الطبقات ، أى نريد إيجاد القيم المثلى لـ (n', n_h) التى تجعلنا نحصل على أقل تباين للتقدير (بتكاليف معينة) وتقدير متوسط المجتمع .
- نعلم أن متوسط المجتمع باستخدام الطبقات يساوى :

$$\mu = \sum_{h=1}^{L} \frac{N_h}{N} \quad \mu_h = \sum_{h=1}^{L} W_h \ \mu_h$$

حيث μ_h هو متوسط الطبقة (h) فى المجتمع و(L) هو عدد طبقات المجتمع .

إن مقدر هذا المتوسط من بيانات العينة الطبقة الثانية التى يتم اختيارها من مفردات العينة الأولى :

$$\overline{x}_{st} = \hat{\mu} = \sum_{h=1}^{L} \frac{n'_h}{n'} \quad \overline{x}_h$$

أى أن :

$$\bar{x}_{st} = \sum_{h=1}^{L} w'_h \, \bar{x}_h \qquad \text{.... (10 - 5)}$$

حيث $w'_h = \frac{n'_h}{n'}$

إن $(\bar{x}_{st})$ هو تقدير غير متحيز لمتوسط المجتمع (μ) .

أما تقدير تباين تقدير متوسط المجتمع فيمكن استخراجه باستخدام الصيغة التالية والذى يعد تقديراً غير متحيز لتباين تقدير المتوسط :

$$V(\bar{x}_{st}) = \sum_{h=1}^{L} \frac{w'^2_h \, s^2_h}{n_h} - \sum_{h=1}^{L} \frac{w'_h \, s^2_h}{N} + \frac{N-n'}{N-1} \, \frac{1}{n'} \sum_{h=1}^{L} w'_h \left(\bar{x}_h - \bar{x}_{st}\right)^2 \qquad \text{.... (10 - 6)}$$

حيث $(\bar{x}_h)$ هو متوسط الطبقة (h) من العينة و(s^2_h) هو تباين الطبقة (h) من العينة أى .

$$\bar{x}_h = \frac{1}{N_h} \sum_{h=1}^{n_h} x_{hi}$$

$$s^2_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(x_{hi} - \bar{x}_h\right)^2$$

ويمكننا استخدام الطريقة نفسها لتقدير نسبة المجتمع فى المعاينة المزدوجة . إن مقدر نسبة المجتمع :

$$\hat{P}_{st} = p_{st} = \sum_{h=1}^{L} w'_h \, p_h \qquad \text{.... (10 - 7)}$$

ومقدر تباين تقدير النسبة :

$$\hat{V}(\hat{P}_{st}) = \sum_{i=1}^{L} w'^2_h \frac{p_h q_h}{n_h - 1} - \frac{1}{N} \sum_{h=1}^{L} w'_h \frac{n_h p_h q_h}{n_h - 1} + \frac{N - n'}{N - 1} \frac{1}{n'} \sum_{h=1}^{L} w'_h (p_h - p_{st})^2$$

(10 - 8) . . . حيث $q_h = 1 - p_h$.

و(p_h) هى نسبة الذين يتصفون بخاصية معينة فى الطبقة (h) من العينة * .

تطبيق (١٠ – ٣)

أوجد تقدير متوسط سنوات الخبرة لموظفى إحدى الوزارات إذا تم تقسيم الموظفين إلى ثلاث طبقات حسب سنوات الخبرة حيث تم اختيار عينة حجمها (٢٠٠) موظف من (١٠٠٠) موظف لتقدير نسبة موظفى كل طبقة (w'_h) ومن ثم سحبت عينة حجمها (١٠٠) موظف من العينة المختارة لتقدير متوسط سنوات الخبرة . وتبين لنا من بيانات العينتين مايلى :

الطبقة	حجم العينة الأولى (n'_h)	حجم العينة الثانية (n_h)	w'_h	$\overline{x}_h$	S^2_h
1	124	62	0.62	3	40.0
2	50	25	0.25	8	30.0
3	26	13	0.13	13	25.0
Total	200	100			

أوجد تقدير متوسط سنوات الخبرة للموظف بمستوى معنوية $(\alpha = 0.05)$

الحـل :

نستخدم الصيغة (4 - 10) لتقدير متوسط سنوات الخبرة :

$$\overline{x}_{st} = \sum_{h=1}^{L} w'_h \, \overline{x}_h$$

$$= (0.62 \times 3) + (0.25 \times 8) + (0.13 \times 13) = 5.55 \text{ (years)}$$

* للحصول على تفاصيل أكثر حول المعاينة المزدوجة يمكن الرجوع إلى Cochran W. G. Sampling techniques. (PP. 327 - 358)

ولتقدير التباين نستخدم الصيغة (6 - 10) فنجد أن :

الحد الأول يساوى :

$$\sum_{h=1}^{L} \frac{w_h^2 s_h^2}{n_h} = \left[\frac{(0.62)^2 \times 40}{62} + \frac{(0.25)^2 \times 30}{25} + \frac{(0.13)^2 \times 25}{13}\right]$$

$$= 0.248 + 0.075 + 0.0325 = 0.3555$$

الحد الثانى يساوى :

$$\sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N}$$

$$= \frac{1}{1000}\left[(0.62 X 40) + (0.25 X 30) + (0.13 X 25) \right]$$

$$= \frac{1}{1000} (24.8 + 7.5 + 3.25) = 0.035$$

الحد الثالث يساوى :

$$\frac{N - n}{N - 1} \quad \frac{1}{n} \sum_{h=1}^{L} w_h \left(\overline{x}_h - \overline{x}_{st}\right)^2$$

$$= \frac{1000 - 200}{1000 - 1} \times \frac{1}{200} \left[0.62(3-5.55)^2 + 0.25 (8-5.55)^2 + 0.13 (13 - 5.55)^2 \right]$$

$$= 0.004 [4.03 + 1.50 + 7.22] = 0.051$$

ويكون تقدير التباين :

$$\hat{V} (\overline{x}_{st}) = 0.3555 - 0.0356 + 0.051 = 0.371$$

أما حدا الثقة لتقدير متوسط المجتمع بمستوى ثقة (٩٥٪) .

$$\overline{x}_{st} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V} (\overline{x}_{st})} = 5.55 \mp 1.96 \sqrt{0.371} = 5.55 \mp 1.19$$

أى أن متوسط سنوات الخبرة يتراوح بين (٤,٣٦) و(٦,٧٤) سنة ، وذلك بدرجة ثقة (٩٥٪) أى أن :

$$4.36 \leqslant \mu \leqslant 6.74$$

تطبيق (١٠ -٤)

باستخدام بيانات التطبيق (١٠ - ٢) ماهو تقدير نسبة المدخنين إذا كان عدد المدخنين فى الطبقات الثلاث كما يلى (a_i ترمز إلى عدد المدخنين فى الطبقة (i)) .

$$a_1 = 15 \ , \ a_2 = 8 \ , \ a_3 = 5$$

الحـل

إن نسبة المدخنين فى الطبقة (i)

$$p_i = \frac{a_i}{n_i}$$

ويكون

$$p_1 = \frac{15}{62} = 0.24 \ , \quad p_2 = \frac{8}{25} = 0.32 \ , \ p_3 = \frac{5}{13} = 0.38$$

ويكون تقدير نسبة المدخنين

$$\hat{P}_{st} = \sum_{h=1}^{L} w'_h p_h$$

$$= (0.62 \text{ X } 0.24) + (0.25 \text{ X } 0.32) + (0.13 \text{ X } 0.38)$$

$$= 0.1488 + 0.08 + 0.0494 = 0.2782$$

أى ٢٧,٨٢٪

ولتقدير التباين نستخدم الصيغة (8 - 10) ويكون :

الحد الأول :

$$\sum_{h=1}^{L} \frac{w'^2_h \, p_h \, q_h}{n_h - 1} = \frac{[(0.62)^2 \, (0.24) \, (1 - 0.24)}{62 - 1} + \frac{(0.25)^2 \, (0.32) \, (1 - 0.32)}{25 - 1}$$

$$+ \frac{(0.13)^2 \, (0.38) \, (1 - 0.38)]}{13 - 1} = 0.00115 + 0.00057 + 0.00033 = 0.00205$$

ويساوى الحد الثانى 0.000204 والحد الثالث 0.00093 ونجد أن التباين يساوى :

$$\hat{V}(p_{st}) = 0.00205 - 0.000204 + 0.00093 = 0.00278$$

ويكون حدا الثقة لتقدير نسبة المجتمع بمستوى ثقة (٩٥٪) .

$$P_{st} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(P_{st})}$$

$$0.2782 \mp 1.96\sqrt{0.00278}$$

$$= 0.2787 \mp 0.1033$$

أى أن

$$0.1752 \leq P_{st} \leq 0.382$$

إن نسبة المدخنين لموظفى هذه الوزارة تتراوح بين (١٧,٥٢٪) و(٣٨,١٢٪) وذلك بمستوى ثقة (٩٥٪) .

١٠ - ٢ المعاينات المتكررة فى مناسبات متعاقبة
(Repeated Sampling on Successive Occasions)

درسنا فيما سبق أنواع المعاينات التى تهدف إلى تقدير معالم ظاهرة ما فى لحظة زمنية معينة ، وهذا النوع من المعاينات يمكن تسميته معاينة فى مناسبة واحدة (Sampling in one Occasion) . ونواجه فى الحياة العملية كثيراً من الحالات التى تتطلب دراسة التغيرات التى تحدث لمعالم المجتمع خلال فترتين زمنيتن أو خلال عدة فترات أو مناسبات متعاقبة ، لذا نقوم بمعاينة المجتمع عدة مرات وفى عدة مناسبات ، وتسمى المعاينة فى هذه الحالة Sampling on Successive Occasions .

ويمكننا توضيح هذا النوع من المعاينات بالبحوث التى تنفذها الأجهزة الإحصائية فى كثير من الدول المتعلقة بتقديرات السكان (العينة السكانية) . إذ كما هو معلوم تقوم هذه الأجهزة بتنفيذ التعداد العام للسكان والمساكن فى فترات زمنية متباعدة (كل خمس أو عشر سنوات) ، ومن الضرورى معرفة التغير الذى يحدث خلال الفترة التى تقع بين تعدادين خاصة وأن المجتمع السكانى يتعرض لتغيرات كثيرة وسريعة . لذا نجد أنه من الضرورى إجراء المعاينة حيث نستعين بسلسلة من العينات الصغيرة (سنويا أو فى فترات أصغر) للحصول على معلومات حديثة . وعندما نكرر إجراء المعاينة ، فإننا نحصل على المعاينات المتكررة (Repeated Samplings) .

ويمكننا التميز بين النوعين التاليين للمعاينات المتكررة :

- المعاينات المتكررة فى مناسبات متعاقبة .
- المعاينة فى مناسبتين .

وسنقوم بدراسة هذين النوعين من المعاينات ، مع التركيز على النوع الثانى .

١٠-٢-١ المعاينات المتكررة فى أكثر من مناسبتين

(Repeated Samplings on Successive Occasions)

عندما يقوم الباحث بمعاينة المجتمع عدة مرات ، يستطيع الحصول على تقديرات حقيقية للتكاليف والتباينات التى يتمكن بواسطتها من استخدام الأساليب الإحصائية المثلى للحصول على تقديرات ذات كفاءة عالية لمعالم المجتمع .

إن تنفيذ سلسلة من المعاينات ، تمكننا من تقديرثلاثة أنوا ع من المقاييس :

١ - التغير فى متوسط المجتمع $(\overline{X})$ الذى يحدث من مناسبة لأخرى وتقديره .

٢ - القيمة المتوسطة لـ $(\overline{X})$ خلال جميع المناسبات وتقديرها .

٣ - متوسط المجتمع $(\overline{X})$ فى الفترة الأخيرة (أحدث مناسبة) وتقديره . ويتوقف اختيار التقدير المناسب على طبيعة المجتمع الذى نقوم بدراسته . مثلاً عندما نرغب فى دراسة وتحديد العوامل التى تتحكم بالمجتمع ، نختار التقدير الأول . وعندما نرغب فى دراسة المجتمعات ذات التغير البطىء خلال عدة فترات زمنية ، نقوم بتقدير القيمة المتوسطة لعدة عينات تنفذ خلال عدة فترات (التقدير الثانى) . كما نقوم بتقدير متوسط المجتمع فى أحدث مناسبة عندما يكون المجتمع سريع التغير .

ولابد لنا من تحديد وحدات العينة عند استخدام المعاينات المتكررة والتى تكون إحدى الحالات التالية :

١ - الاحتفاظ بنفس الوحدات للحصول على بيانات فى المناسبات المختلفة أى نستخدم الوحدات نفسها فى كل مناسبة اختيار .

٢ - اختيار وحدات جديدة فى كل مناسبة .

٣ - الاحتفاظ بجزء من العينة الأولى واستبدال الجزء الآخر بوحدات جديدة فى المناسبة الأخرى .

إن اختيار أى من هذه الحالات يتوقف على تباين الوحدات ، وتباين التغير فى المتوسطات ، والأهمية النسبية للمعلومات المطلوب جمعها . وهكذا يمكننا القول إنه عندما نرغب فى تقدير التغير ، يُفضل استخدام وحدات العينة السابقة نفسها فى جميع المناسبات . أما عندما نريد تقدير كل متوسط خلال جميع المناسبات ، فيفضل اختيار عينة جديدة فى كل مناسبة . أما عندما نرغب فى الحصول على تقديرات للمناسبة الحالية ، فنحصل على الدقة نفسها ، إما باستخدام وحدات العينة نفسها أو بتبديلها فى كل مناسبة . ولكن البديل الأفضل هو استبدال جزء من العينة والإبقاء على جزء منها فى كل مناسبة . ولابد لنا من الإشارة إلى أنه عندما نستخدم وحدات العينة نفسها فى عدة مناسبات ، قد ينتج بعض التحيز إذ قد تحدث بعض التغيرات نتيجة للملاحظات التى تلقاها المدلى بالبيانات عند جمع البيانات فى الزيارة الأولى .

إن الحصول على تقديرات معالم المجتمع وتبايناتها للمعاينة فى عدة مناسبات معقد إلى حد ما ، لذا سنركز على التقديرات التى نحصل عليها من المعاينة فى مناسبتين متعاقبتين :

١٠-٢-٢ المعاينة فى مناسبتين متعاقبتين :
(Sampling on Two Successive Occasions)

لإيجاد تقدير متوسط المجتمع باستخدام المعاينة فـى مناسبتين متعاقبتين ، نستخدم إحدى الطريقتين التاليتين حسب الوحدات التى تحتويها العينة الثانية :

- إذا أخذنا عينتين مستقلتين فى المناسبتين أو استخدمنا وحدات العينة نفسها فى كل من المناسبتين (الحالتان الأولى والثانية) ، نعد كل مناسبة منهما مستقلة ، ونوجد التقدير لكل عينة بصرف النظر عن القيم التى حصلنا عليها فى المناسبة الأخرى . وتسمى هذه التقديرات «التقديرات العامة أو الشاملة» (Over all estimates) . ويحتوى التقدير فى هذه الحالة على جميع المعلومات التى تحصل عليها باستخدام إحدى هاتين الحالتين .
- إذا احتفظنا بجزء من وحدات العينة الأولى واستبدلنا الجزء الآخر بوحدات جديدة ، نقوم بتقدير معالم المجتمع بطريقة تمكننا من الحصول على تقدير أفضل بإدخال تقديرى الجزء المحتفظ به والجزء الجديد .

وسنقوم بدراسة الطريقة الثانية ، حيث تم معالجة الطريقة الأولى عند دراسة أنواع العينات إذ تعد العينتان المتعاقبتان مستقلتين ، أى نعالج كل عينة وحدها .

لنفرض أن حجم العينة (n) ثابت فى المناسبتين ونريد تقدير متوسط المجتمع وتباينه فى المناسبة الثانية عن طريق الاحتفاظ بجزء من العينة فى المناسبة الأولى واستبدال الجزء الآخر بوحدات جديدة مع ثبات حجم العينة فى كلتا المناسبتين (n) ، وذلك لتبسيط العمليات الرياضية ، لنستخدم الرموز التالية :

n حجم العينة فى كلتا المناسبتين .

$\overline{x}_1$ متوسط العينة فى المناسبة الأولى .

$\overline{x}_2$ متوسط العينة فى المناسبة الثانية .

m عدد الوحدات المحتفظ بها فى العينة الثانية من العينة الأولى .

u عدد الوحدات التى سيتم استبدالها $(u=n-m)$.

فى العينة الثانية .

$\overline{x}_{1m}$ متوسط العينة للجزء المحتفظ به من العينة الأولى .

$\overline{x}_{1u}$ متوسط العينة للجزء الذى سيستبدل من العينة الأولى .

$\overline{x}_{2m}$ متوسط العينة للجزء المحتفظ به للعينة الثانية .

$\overline{x}_{2u}$ متوسط العينة للجزء الجديد للعينة الثانية .

ويمكننا الحصول على تقدير متوسط المجتمع باستخدام تقدير متوسط المجتمع فى الجزء المحتفظ به وتقديره فى الجزء الجديد ، وذلك من بيانات العينة الثانية (أى العينة فى المناسبة الثانية) أى نريد استخراج قيمة $(\overline{x}_{2m})$ وقيمة $(\overline{x}_{2u})$ وتقدير متوسط المجتمع من هاتين القيمتين باستخراج الوسط المرجح لهما وأن يكون الترجيح بمعكوس ثباتيهما وفقًا للخطوات التالية :

١ – نقوم بتقدير متوسط العينة للوحدات الجديدة من بيانات العينة فى المناسبة الثانية $(\overline{x}'_{2u})$ وذلك باستخدام الصيغة :

$$\overline{x}'_{2u} = \frac{1}{u} \sum_{i=1}^{u} \overline{x}_{2ui} = \overline{x}_{2u} \quad .... (10-9) .$$

حيث (x_{2ui}) تمثل مفردات العينة فى المناسبة الثابتة الجديدة ويمكن الحصول على تقدير الوحدات المحتفظ بها فى المناسبة الأولى وبياناته من المناسبة الثابتة

باستخدام طريقة التقدير بالانحدار فى المعاينة المزدوجة ولنرمزله بالرمز $(\overline{x}'_{2m})$:

$$\overline{x}'_{2m} = \overline{x}_{2m} + b\,(\overline{x}_1 - \overline{x}_{1m}) \quad \text{.... (10 - 10)}$$

٢ - نستخرج تباين المتوسط من الجزء الجديد (الوحدات الجديدة فى العينة الثانية)

باستخدام الصيغة $V(\overline{x}'_{2u}) = \frac{S_2^2}{u}$ ولنرمز له بالرمز $\left(\frac{1}{w_{2u}}\right)$.

أما تباين المتوسط $(\overline{x}'_{2m})$ أى التباين من بيانات المناسبة الثانية باستخدام طريقة الانحدار فيساوى :

$$V(\overline{x}'_{2m}) = \frac{S_2^2 \quad (1-r)^2}{m} + r^2\,\frac{S_2^2}{n} \quad \text{.... (10 - 11)}$$

ولنرمز له بالرمز $\left(\frac{1}{w_{2m}}\right)$ وذلك بإهمال معامل تصحيح المجتمع المحدود عندما يكون حجم المجتمع كبيراً حيث (r) هو معامل الارتباط بين أزواج البيانات (x_{1m}, x_{2m}) .

٣ - إن أفضل تقدير لـ $(\overline{X}_2)$ يمكن الحصول عليه بترجيح التقديرين المستقلين بمعكوس تباينهما . فإذا كان (w_{2m}, w_{2u}) هما معكوس تباينهما فإن أفضل تقدير لـ $(\overline{X}_2)$ هو :

$$\overline{x}'_2 = \phi_2\,\overline{x}'_{2u} + (1 - \phi_2)\,\overline{x}'_{2m} \quad \text{.... (10 - 12)}$$

حيث

$$\phi_2 = \frac{w_{2u}}{w_{2u} + w_{2m}}$$

وباستخدام طريقة المربعات الصغرى ، نجد أن :

$$V(\overline{x}'_2) = \frac{1}{w_{2u} + w_{2m}}$$

أى يساوى :

$$V(\overline{x}'_2) = \frac{S^2_2 \quad (n - ur^2)}{n^2 - u^2 r^2}$$

وللحصول على القيمة المثلى للتباين نأخذ تفاضل $V(\overline{x}'_2)$ بالنسبة للاختلاف فى (u) وهذا يعطى :

$$\frac{u}{n} = \frac{1}{1 + \sqrt{1 - r^2}}$$

$$\frac{m}{n} = \frac{\sqrt{1 - r^2}}{1 + \sqrt{1 - r^2}} \quad .... (10 - 13)$$

وعند تبديل (u) المثلى فى $V(\overline{x}'_2)$ نجد أن التباين الأمثل يساوى :

$$Vopt\ (\overline{x}'_2) = \frac{S^2}{2n}\ (1 + \sqrt{1 - r^2}) \quad .... (10 - 14)$$

إن الصيغة (13 - 10) تمكننا من تحديد النسبة المئوية من الحجم الأمثل الذى يجب الاحتفاظ به إلى المناسبة الثانية باستخدام $\frac{m}{n}$ وذلك حسب قيم معامل الارتباط (r) .

ويمكننا استخراج الزيادة النسبية فى الدقة التى نحصل عليها نتيجة اختيار عينة جديدة فى المناسبة الثانية بمقارنة تباين متوسط العينة مع التباين الأمثل أى تكون الزيادة النسبية فى الدقة ونرمز لها بالرمز ($\triangle$) وتساوى :

$$\triangle = \frac{s^2 / n}{Vopt\ (\overline{x}'_2)} - 1 = \frac{2}{1 + \sqrt{1 - r^2}} - 1 \quad .... (10 - 15)$$

وعندما r = 1 لانحتفظ بأية وحدات إلى المناسبة الثانية وعندما تكون (r) غير معلومة ، يمكن تقديرها من بيانات العينة أو من معلومات سابقة *

* لمزيد من التفاصيل ، راجع Cochran W. G. : Sampling Techniques, pp. (346 - 348)

تطبيق (١٠ – ٥)

ماهى النسبة المئوية للحجم الأمثل الذى يجب الاحتفاظ به إلى المناسبة الثانية والزيادة النسبية فى الدقة التى نحصل عليها إذا كان $(r = 0.80)$.

الحل

$$\frac{m}{n} = \frac{\sqrt{1 - r^2}}{1 + \sqrt{1 - r^2}} = \frac{\sqrt{1 - 0.80^2}}{1 + \sqrt{1 - 0.80^2}} = 0.38$$

وتكون الزيادة النسبية فى الدقة $\triangle$ مساوية إلى

$$\triangle = \frac{2}{1 + \sqrt{1 - r^2}} - 1 = \frac{2}{1 + \sqrt{1 - 0.80^2}} - 1 = \frac{2}{1.6} - 1 = 0.25$$

أى أن الزيادة فى الدقة التى نحصل عليها هى 0.25 أى ٢٥٪ .

تطبيق (١٠ – ٦)

تمثل البيانات التالية عدد المعاملات التى أنجزها (١٢) موظفًا خلال شهرى رجب وشعبان . تبين من دراسة سابقة أن معامل الارتباط بين أزواج العينتين $s^2_2 = 50 , r = 0.90$

رقم الموظف	عدد معاملات رجب	عدد معاملات شعبان	رقم الموظف	عدد معاملات رجب	عدد معاملات شعبان
١	٢٠	–	٧	٢٤	٢٣
٢	٢٢	–	٨	٢٣	٢٢
٣	٢٤	–	٩	–	٢٤
٤	٢٣	–	١٠	–	٢٢
٥	٢٥	٢٤	١١	–	٢١
٦	١٩	٢٥	١٢	–	٢٢

أوجد

١ – تقدير متوسط عدد المعاملات التي ينجزها الموظف .

٢ – تباين تقدير المتوسط .

الحل

– نستخرج التقديرات التالية :

$$\bar{x}_{2u} = \frac{1}{4}(24 + 22 + 21 + 22) = 22.25$$

$$\bar{x}_{1} = \frac{1}{8}(20 + 22 + ... + 23) = 22.5$$

$$\bar{x}_{1m} = \frac{1}{4}(25 + 19 + 24 + 23) = 22.75$$

$$\bar{x}_{2m} = \frac{1}{4}(24 + 25 + 23 + 22) = 23.5$$

– نحسب معامل الانحدار $(\hat{B})$ المقدر من بيانات عدد المعاملات المنجزة خلال الشهرين من المشتركة (للموظفين ٥ ، ٦ ، ٧ ، ٨) :

فنجد أن $\hat{B} = -0.265$

– نستخرج التقديرات التالية :

$$\frac{s^2}{n} = \frac{50}{12} = 4.16$$

$$w_u = \frac{12}{50} = \frac{1}{4.16} = 0.24$$

والتباين الأمثل (أقل تباين) المقدر :

$$V(\overline{x}'_{2m}) = \frac{s_2^2\ (1 - r^2)}{m} + r^2\ \frac{s_2^2}{n}$$

$$= \frac{50\ (1 - 0.90^2)}{8} + (0.90)^2\ \frac{50}{12}$$

$$= 1.188 + 3.375 = 4.563$$

$$\frac{1}{w_m} = \frac{1}{4.563} = 0.219$$

$$\phi_2 = \frac{w_u}{w_u + w_m} = \frac{0.24}{0.459} - 0.523$$

ويكون تقدير المتوسط :

$$\overline{x}'_2 = \phi_2\ \overline{x}_{2u} + (1 + \phi_2)\ \overline{x}'_{2m}$$

$$= 0.523 \times 22.25 + (1 - 0.523)\ \overline{x}'_{2m}$$

$$\overline{x}'_{2m} = \overline{x}_{2m} + \hat{\beta}\ (\overline{x}_1 - \overline{x}_{1m})$$

$$= 23.5 + (-0.265)\ (22.5 - 22.75) = 23.57$$

$$\overline{x}'_2 = (0.523 \times 22.25) + (0.477 \times 23.57$$

$$= 11.64 + 11.25 = 22.89$$

أما تباين تقدير المتوسط فيساوى :

$$V(\overline{x}'_2) = \frac{1}{w_{2u} + w_{2m}}$$

$$= \frac{1}{0.24 + 4.563} = \frac{1}{4.803} = 0.21$$

١٠ - ٢ المعاينة المساحية (Area Sampling)

تعد المعاينة المساحية من المعاينات التي تستخدم بشكل واسع ، خاصة إذا كانت وحدات المعاينة هي المساكن أو الأفراد والأسر أو المزارع أو المخازن ، ونستطيع بواسطة هذا النوع من المعاينات ، تكوين أطر متعددة لكثير من البحوث ، خاصة تلك التي تتعلق بالمساكن والأسر عندما تكون مجتمعاتها كبيرة ولايتوافر عنها إطارات حديثة أو يتطلب إعدادها نفقات كبيرة .

ونستطيع تلخيص الخطوات الواجب اتباعها لاختيار وحدات العينة المساحية بما يلي :

- إعداد الخرائط التي تتضمن الحدود الجغرافية للمناطق التي يشملها البحث بمقياس كبير .
- تقسيم المساحة الكلية إلى مناطق جغرافية رئيسية (طبقات) بحيث تحتوى كل منطقة على عدد من القطاعات الرئيسية ، (Blocks) .
- ترقيم القطاعات بأرقام مسلسلة .
- يحتوى كل قطاع على عدد من القطاعات الفرعية (Segments) ويتم ترقيمها بأرقام مسلسلة .
- تحديد نوع العينة التي ستستخدم وحجمها . فإذا تقرر استخدام العينة الطبقية ، فإننا نختار عدداً من القطاعات الفرعية عشوائيًا من كل قطاع رئيسي من جميع القطاعات المكونة للمجتمع . أما إذا استخدمت المعاينة العنقودية ذات المرحلتين ، فإننا نختار عدداً من القطاعات الرئيسية كمرحلة أولى ، ومن ثم نختار عدداً من القطاعات الفرعية من كل قطاع رئيسي تم اختياره ، وقد تستخدم المعاينة ذات المراحل المتعددة إذا كانت القطاعات الفرعية المقسمة إلى أجزاء صغيرة .
- يتم اختيار الوحدات من الإطارات التي يتم تكوينها لتوضيح هذه الخطوات ، نفترض أننا نرغب في اختيار عدد من الأسر من إحدى المدن . نقوم بتحضير خارطة لهذه المدينة بمقياس كبير ونقسمها إلى قطاعات رئيسية يمثل كل قطاع مساحة معينة ، ويتم تقسيم هذه القطاعات إلى طبقات تمثل كل طبقة (حيًا) . ويتم تحديد حدود كل من هذه الأحياء (الطبقات) . يقسم كل حي إلى قطاعات فرعية ، يمثل كل منها مساحة معينة ويتكون من عدد من المباني .

إذا استخدمنا المعاينة العنقودية الطبقية فإننا نختار من كل طبقة عدداً من القطاعات الرئيسية عشوائيًا ، ثم نختار عدداً من المباني من كل قطاع تم اختياره (كمرحلة ثانية) . ويمكننا اختيار عدد من الأسر من المباني المختارة ، ونلاحظ أننا نستطيع تكوين إطارات للأحياء والقطاعات الفرعية والمباني والأسر ، ويساعد ذلك على تنفيذ البحوث التي وحداتها الأسر أو المباني أو المساحات (كالمزارع) .

وتستخدم المعاينة المساحية عندما نرغب في معاينة إحدى الغابات أو المناطق الزراعية ، حيث تؤخذ للغابة صور فوتوغرافية من الجو لتقسيمها إلى طبقات حسب كثافة عدد الأشجار وأنواعها ويتم اختيار عدد من المساحات من كل طبقة ويتم إجراء الدراسة عليها . كذلك تستخدم المعاينة المساحية عندما يرغب الباحث في تقدير معالم المجتمع لمجتمعات حركية تتنقل من مكان لآخر كالأسماك في البحار والأنهار والطيور ، حيث توجد صعوبة في حصرها حصراً شاملاً ، مثلاً لتقدير كمية الأسماك في منطقة ما ، نقسم هذه المنطقة إلى مساحات يتم تصنيفها في طبقات حسب معايير محددة . ويتم اختيار عدد من المساحات عشوائيًا ، وتصطاد الأسماك المتواجدة فيها ، ويمكننا تقدير عدد الأغنام والحيوانات في الغابات بالطريقة نفسها وباستخدام الطرق الإحصائية المناسبة المستخدمة في المعاينة الطبقية أو المعاينات الأخرى .

١٠ - ٤ المعاينة من المجتمعات البرية :

(Sampling from Wildlife Populations)

يوجد أنواع أخرى من المعاينات تستخدم لمعاينة المجتمعات البرية ، لدراسة نموها والمحافظة عليها وتقدير أعدادها . ويهتم هذا النوع من المعاينات بتقدير حجم المجتمع (N) للمجتمعات البرية الكبيرة .

وتستخدم طريقتان لتقدير حجم المجتمع :

١٠-٤-١ الطريقة المباشرة لمعاينة المجتمعات البرية : (Direct Method)

تعتمد هذه الطريقة على اختيار عينة عشوائية من وحدات المجتمع الذي ندرسة ، ونقوم بوضع علامات مميزة على كل وحدة من وحدات العينة المختارة ، ونعيدها إلى مجتمعاتها . ونقوم في وقت لاحق باختيار عينة عشوائية أخرى ذات حجم محدد من وحدات المجتمع نفسه ، ويتم حصر عدد الوحدات التي تحمل العلامات التي تم وضعها ، ونقوم بتقدير نسبة الوحدات التي تحمل علامات ، ومن ثم تقدير حجم المجتمع .

نفترض أن (N) يمثل حجم المجتمع أي عدد الحيوانات التي نرغب في معاينتها و(t) يمثل عدد الوحداب التي تم وضع علامات عليها . إن نسبة الوحدات ذات العلامات إلى إجمالي المجتمع يساوي عدد الحيوانات التي تم تعليمها إلى إجمالي عدد الحيوانات ، أي

$$P = \frac{t}{N} \quad .... (10 - 16)$$

ويمكننا إيجاد حجم المجتمع أى عدد الحيوانات من الصيغة :

$$N = \frac{t}{P} \quad .... (10 - 17)$$

ونستطيع تقدير حجم المجتمع (N) حيث نستطيع تقدير (P) من العينة التى يتم اختيارها حيث (t) معلومة باستخدام الصيغة التالية :

$$\hat{N} = \frac{t}{\hat{P}} \quad .... (10 - 18)$$

حيث ($\hat{P}$) تمثل نسبة الوحدات المعلمة فى العينة الثانية ، ويتم استخراجها باستخدام الصيغة التالية :

$$\hat{P} = \frac{s}{n} \quad .... (10 - 19)$$

حيث ترمز (s) إلى عدد الوحدات المعلمة بالعينة و(n) إلى حجم العينة وبالتالى يكون تقدير حجم المجتمع فى هذا النوع من المعاينات :

$$\hat{N} = \frac{t}{\hat{P}} \quad .... (10 - 20)$$

أى يساوى :

$$\hat{N} = \frac{nt}{s} \quad .... (10 - 21)$$

أما تقدير تباين (N) فيساوى :

$$\hat{V}(\hat{N}) = \frac{t^2 \ n (n - s)}{s^3} \quad .... (10 - 22)$$

ويمكننا استخراج حدى الثقة باستخدام الصيغة :

$$\hat{N} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V}(\hat{N})} \quad .... (10 - 23)$$

حيث (Z) تمثل القيمة الجدولية للتوزيع الطبيعى بمستوى ثقة % $(1 - \alpha)$.

١٠-٤-٢ طريقة المعاينة العكسية من المجتمعات البرية
(Inverse Sampling Method)

يتمثل الاختلاف الرئيسى بين الطريقة المباشرة والطريقة العكسية للمعاينة من المجتمعات البرية فى أن حجم العينة فى الطريقة العكسية يكون غير محدد ، حيث يتم اختيار الوحدات حتى نحصل على عدد محدد من الوحدات ذات العلامات التى تم وضعها على الوحدات :

لنفرض أن حجم المجتمع الذى نريد معاينته (N) وأن حجم العينة الأولى التى تم اختيارها (t) تم وضع علامات مميزة عليها وتم إعادتها . وبعد فترة يتم اختيار عينة عشوائية حجمها (n) وحدة ويكون تقدير نسبة الوحدات المعلمة التى عددها(s) : $\hat{p} = \frac{s}{n}$

ويكون تقدير حجم المجتمع باستخدام الصيغة ($\hat{N} = \frac{t}{\hat{p}}$) أى أن :

$$\hat{N} = \frac{nt}{s} \quad .... (10 - 24)$$

أما تقدير تباين تقدير حجم المجتمع فيساوى :

$$\hat{V}(\hat{N}) = \frac{t^2 \ n (n - s)}{s^2 \ (S + 1)} \quad .... (10 - 25)$$

ونقوم بتقدير حدى الثقة باستخدام الصيغة :

$$\boxed{\hat{N} \mp Z\sqrt{\hat{V}(\hat{N})}} \quad .... (10 - 26)$$

حيث (Z) تمثل القيمة الجدولية للتوزيع الطبيعي بمستوى ثقة % $(1-\alpha)$. ولابد لنا من الإشارة إلى أن الطريقة العكسية للمعاينة تعطى معلومات أكثر دقة من الطريقة المباشرة حيث تعطى الطريقة الثانية حجم العينة (n) الذى يضمن الحصول على (s) وحدة معلمة لتقدير حجم المجتمع . ويتم تحديد حجم العينة إذا عرفنا حجماً تقريبياً للمجتمع (N) حيث نستطيع تحديد التباين $V(\hat{N})$ لأحجام معينة من العينات (t) و(n) ويتم تحديد هذه الأحجام بالتعبير عنها بكسر من (N) * .

تطبيق (١٠ - ١)

ترغب إحدى الهيئات تقدير حجم مجتمع الحبارى فى منطقة ما لتنظيم موسم الصيد . سحبت عينة من الحبارى حجمها $(t = 300)$ حيث وضعت عليها علامات مميزة وأعيدت إلى المنطقة التى تعيش فيها . وبعد مرور شهر تم اختيار عينة حجمها $(n = 200)$ وجد منها $(s = 75)$ تحمل العلامات التى تم وضعها . ماهو تقدير عدد الحبارى بمستوى ثقة (٩٥٪) .

الحل :

نعلم أن

$$\hat{N} = \frac{nt}{s}$$

$$= \frac{200 \times 300}{75} = 800$$

$$\hat{V}(\hat{N}) = \frac{t^2 n (n - s)}{s^3}$$

$$= \frac{(300)^2 \times (200)(200 - 75)}{(75)^3} = 5333.33$$

---

* لمزيد من التفاصيل ، راجع .

Scheafer, Mendenhall and Ott Elementary Survey Sampling, Duxbury Press, 1979 (pp. 221 - 225)

ويكون حدا الثقة لتقدير حجم المجتمع :

$$\hat{N} \mp Z\sqrt{\hat{V}(\hat{N})}$$

$$800 \mp 1.96\sqrt{5333.33}$$

$$= 800 \mp 143$$

أى أن عدد الحبارى فى المنطقة يتراوح بين (٦٥٧) و (٩٤٣) بمستوى ثقة ٩٥٪
أى أن :

$$657 \leq N \leq 943$$

تطبيق (١٠ - ٢)

لتقدير عدد الطيور فى إحدى المناطق تم اختيار عينة عشوائية حجمها $(t = 400)$ من الطير تم وضع حلقات فى أرجلها وإعادتها للمنطقة ، وبعد شهر تم اختيار عينة من الطيور حتى أصبح عدد الطيور التى تحمل الحلقات $(s = 100)$ وكان حجم العينة الثانية $(n = 300)$.
ماهو تقدير حجم المجتمع بمستوى ثقة ٩٥٪ .

الحـل

$$\hat{N} = \frac{nt}{s}$$

$$= \frac{300 \times 400}{100} = 1200$$

$$\hat{V}(\hat{N}) = \frac{t^2 n (n - s)}{s^3 (s + 1)}$$

$$= \frac{(400)^2 \times (300)(300 - 100)}{(100)^2 (100 + 1)} = 9505$$

ويكون حدا الثقة :

$$\hat{N} \mp Z \sqrt{\hat{V}(\hat{N})}$$

$$1200 \mp 1.96 \sqrt{9505}$$

$$1200 \mp 191$$

ويكون حدا الثقة (١٠٠٩) و(١٣٩١) أى أن عدد الطيور يتراوح بين (١٠٠٩) و(١٣٩١) طيرًا بمستوى ثقة ٩٥٪ أى أن :

$$1009 \leq N \leq 1391$$

الفصل
الحادى عشر

استخدام الحاسوب فى مجال الميـ[illegible]

١١ - ١ تمهيد

تطورت تقنيات الحاسوب فى السنوات الأخيرة تطوراً سريعاً وتعددت مجالات استخدامه لتشمل كافة المجالات الاقتصادية والاجتماعية . ويعد استخدام الحاسوب في مجال البحوث من أهم الاستخدامات التى أدت إلى تطور سريع فى إنجازها بسبب السرعة والدقة التى يتصف بها الحاسوب خاصة عند إنجاز العمليات الرياضية المعقدة وتبويب البيانات. واستخراج أهم المقاييس الإحصائية وتحليل البيانات .

لقد اهتم الباحثون بالحاسوب عند تنفيذ بحوثهم خاصة عند استخدام أسلوب المعاينة كأسلوب لجمع البيانات وذلك لاختيار وحدات العينة وعرض البيانات جدولياً وبيانياً وتقدير معالم المجتمع . كما استخدم الحاسوب لتحليل البيانات بالدقة والسرعة الفائقة .

١١ - ٢ البرامج الإحصائية الجاهزة

لقد تعددت لغات البرمجة المستخدمة فى الحاسوب وتطورت تطوراً كبيراً فسايرت التطور الذى حدث فى تقنياتها ومجالات استخدامها . لقد كانت لغات البرمجة الفورتران FORTRAN والكوبول COBOL والبيسك BASIC وأسمبلى ASSEMBLY وبى إل/واحد PL/1 وغيرها من لغات البرمجة ، اللغات الأساسية التى استخدمها الباحثون لتنفيذ بحوثهم وقامت الشركات المتخصصة بلغات البرمجة وتقنيات الحاسوب بإعداد أنظمة جاهزة متعددة لتسهيل تنفيذ البحوث والقيام بالعمليات التى يحتاجها الباحثون بالسرعة والدقة المطلوبة .

وتعد أنظمة MINITAB وSAS وSPSS من أهم هذه الأنظمة التى تستخدم للأغراض الإحصائية . وسنقوم باستعراض نظام MINITAB بسهولة استخدامه في الحاسوب الشخصى ونظام SAS ونظام SPSS لانتشار استخدامها فى الحواسيب الضخمة والشخصية أيضاً .

١١-٢-١ نظام MINITAB

لقد صمم هذا النظام فى عام ١٩٧٢م للمهتمين بدراسة المواد الإحصائية ، ثم طور ليخدم المتخصصين فى مجالات الهندسة والعلوم الاجتماعية والنفسية والإدارية والعلوم الأخرى* ويتصف هذا النظام بكونه سهل الاستخدام فى مجال العينات خاصة للذين ليس لديهم خبرة سابقة في مجال الحاسوب ويحتوى هذا النظام على إمكانيات كبيرة تساعد الباحثين فى تنفيذ

* Ryan & Others . Minitab . Duxbury Press, Poston, 1985 (P.m)

بحوثهم خاصة عند اختيار وحدات العينة وتقدير معالم المجتمع وذلك إضافة للعمليات الإحصائية المتعلقة بعرض البيانات جدولياً وبيانياً وتحليلها وذلك باستخدام الحاسوب الشخصي .

١١-٢-٢ نظام التحليل الإحصائي ساس (SAS) (Statistical Analysis System)

يعد نظام التحليل الإحصائي «ساس» ، من أكثر أنظمة البرامج الجاهزة استخداماً بسبب المرونة والسرعة الفائقة في التعامل مع البيانات وعرض البيانات جدولياً وبيانياً واستخراج أهم المقاييس الإحصائية والقيام بالتحليلات المناسبة والتنبؤ بأهم القيم المستقبلية وكتابة وطباعة التقارير ، سواء كانت التقارير التي تكتب بشكل تلقائي أو التقارير التي يرغب الباحث بطباعتها في أشكال معينة . وقد تم إعداد نظام ساس باستخدام لغات البرمجة الرئيسية ASSEMPLY (حوالي ٥٠٪ من النظام) و PL/1 (حوالي ٤٨٪ من النظام) وFORTRAN (حوالي ٢٪ من النظام) * .

ويعد هذا النظام من أفضل الأنظمة الإحصائية باستخدام الحواسيب الضخمة ، ويستخدم أيضاً في الحاسوب الشخصي .

١١-٢-٣ حقيبة البرامج الإحصائية للعلوم الاجتماعية (SPSS)
(Statistical Package for the Social Sciences)

لقد صممت هذه الحقيبة لتحليل بيانات المسوحات خاصة في مجال العلوم الاجتماعية . وتتصف هذه الحقيبة بإمكانات كبيرة لتكوين الجداول وتحتوي على برامج لاستخراج أهم المقاييس الإحصائية وتحليل البيانات ، (خاصة المتعلقة بالارتباط والانحدار وتحليل التباين والتغاير والتحليل العاملي وغيرها) **

تستخدم الأنظمة الثلاثة السابقة في مجال العينات إذ يعد نظام (Minitab) من أفضل الأنظمة الثلاثة في اختيار وحدات المعاينة عشوائياً كما يعد نظاماً مرناً في تقدير معالم المجتمع سواء كان التقدير بنقطة أو التقدير بفترة .

أما نظام (SAS) فيعد من أفضل الأنظمة في مجال تقدير معالم المجتمع من بيانات عينة خاصة إذا كان حجم العينة كبيراً . ويتصف نظام (SPSS) بسهولة استخدامه في البحوث الاجتماعية خاصة لاستخراج بعض المقاييس الإحصائية .

---

* خالد بالطيور مقدمة في التحليل الإحصائي مع برنامج SAS ، مؤسسة جمال الجاسم للإلكترونيات ، الدمام ١٩٩٠م .

** لمزيد من التفاصيل ، راجع

Yates Frank . Sampling Methods for Censuses and survey. Charles & Company Ltd. 1981 (P 393) .

وسنقوم بالتركيز على نظامى (Minitab) و(SAS) نظراً لاستخدامها فى مجال العينات بشكل واسع .

١١ -٢ استخدام نظام (MINITAB) فى مجال العينات

يستخدم نظام (Minitab) لاختيار وحدات المعاينة عشوائياً وعرض البيانات وتقدير معالم المجتمع من بيانات العينة وسنقوم باستعراض الأوامر المتعلقة بذلك .

١١-٢-١ اختيار وحدات المعاينة عشوائيا :

يمكننا تقسيم الأوامر المتعلقة باختيار الوحدات عشوائياً إلى قسمين رئيسين

١ - أوامر تتعلق بالعينات العشوائية من التوزيعات الإحصائية النظرية كتوزيع ذى الحدين والتوزيع الطبيعى وتوزيع بواسون وغيرها من التوزيعات .

٢ - أوامر تتعلق بالعينات العشوائية من المجتمعات الإحصائية المحدودة الفعلية (Actual Finite Populations) .

وسنركز على الأوامر المتعلقة باختيار وحدات المعاينة من المجتمعات المحدودة الفعلية لأهميتها عند اختيار وحدات العينة فى التطبيقات العملية .

عندما نقوم باختيار عينة عشوائية بسيطة بدون إرجاع من مجتمع محدود نستخدم الأمر التالى :

SAMPLE N observation From C, ... , C
Put into C, ... , C

ويمكننا هذا الأمر من اختيار عينة حجمها (n) وحدة من البيانات فى المجموعة الأولى من الأعمدة وتخزين العينة فى المجموعة الثانية من الأعمدة . والعينة التى تم اختيارها هى عينة عشوائية بسيطة بدون إرجاع .

تطبيق (١١ – ١)

ترغب إحدى الجهات فى اختيار عينة عشوائية حجمها (١٥) موظفاً من موظفيها البالغ عددهم (١٠٠) موظف لتقدير متوسط سنوات الخبرة .

المطلوب تحديد أرقام الموظفين الذين تم اختيارهم .

الحـل :

إن الأرقام المحددة فى (C4) تمثل الأرقام التى تم اختيارها

```
MTB  > Set C3
MTB  > 1 : 100
MTB  > end
MTB  > Print C3
        1    2    3    4    5    6    7    8    9
10      11
        12   13   14   15   16   17   18   19   20
21      22
        23   24   25   26   27   28   29   30   31
32      33
        34   35   36   37   38   39   40   41   42
43      44
        45   46   47   48   49   50   51   52   53
54      55
56      57   58   59   60   61   62   63   64
65      66
        67   68   69   70   71   72   73   74   75
67      77
        78   79   80   81   82   83   84   85   86
87      88
        89   90   91   92   93   94   95   95   97
98      99
        100

MTB  > Sample 15 C3 C4
MTB  > Print C4
C4

        2    83   96   68   65   17   46   73   18   69   53
74      99
        49   59
```

١١-٢-٢ عرض بيانات العينة جدوليا

يستخدم الأمر TABLE لعرض بيانات العينة فى الجدول المناسب :

TABLE the data Classified by C , C, ... C.

مثلاً إذا أردنا تبويب بيانات عينة الموظفين حسب الإدارة التى يعملون بها (Dep.) والمدينة ، تستخدم الأمر :

TABLE 'Dep' 'CITy' ,
TOTPERCENT .

توضح TOTPERCENT النسب المئوية للإجمالى ويمكن أن نضع ROWPERCENT أو COLPERCENT لوضع النسب المئوية للأسطر أو الأعمدة أو لا نضع أيًا منها .

تطبيق (١١ - ٢)

توضح البيانات التالية سنوات الخبرة لـ (١٢) موظفًا (EXP) تم اختيارهم عشوائيًا ، والأقسام التى يعملون بها (Dep) ، والمدن التى يعملون بها (City) .

المطلوب عرض بيانات الموظفين حسب القسم والمدينة :

سنوات الخبرة : ٣ ، ٤ ، ٥ ، ٦ ، ٣ ، ٨ ، ٤ ، ٤ ، ٥ ، ٣ ، ٦ ، ٧ ، ٤

القســـــم : ١ ، ٢ ، ١ ، ١ ، ٢ ، ١ ، ٢ ، ١ ، ٢ ، ٢ ، ٢ ، ٢ ، .

المدينـــــة : ١ ، ٢ ، ٢ ، ١ ، ٢ ، ١ ، ١ ، ١ ، ٢ ، ٢ ، ١ ، ١ ،

الحـل

نكتب الأوامر التالية التى تعطى النتائج الواردة بعدها

```
MTB > read   c1 - c3
MTB > data   > 3  1  1
MTB > data   > 4  2  2
MTB > data   > 5  1  2
MTB > data   > 6  1  1
MTB > data   > 7  2  2
MTB > data   > 8  1  1
.....
....
```

```
MTB >data    > 4   2   1
MTB >data    > end
MTB >name  c1    'exp'
MTB >name  c2    'dep'
MTB >name  c3    'city'

MTB >print   c1 - c3

ROW    exp     dep     city

  1     3       1       1
  2     4       2       2
  3     5       1       2
  4     6       1       1
  5     3       2       2
  6     8       1       1
  7     4       2       1
  8     4       1       1
  9     5       2       2
 10     3       2       2
 11     6       2       1
 12     7       2       1
 13     4       2       1

MTB  > table   c2  c3 :
SUBC > totpercent.

ROWS    dep    COLUMNS    city

          1          2          ALL

  1     30.77       7.69       38.46
  2     30.77      30.77       61.54
 ALL    61.54      38.46      100.00

CELL CONTENTS --
                          % OF TBL
```

١١-٣-٢ تقدير معالم المجتمع

أ - تقدير الوسط الحسابى بنقطة

نستخدم الأمر DESCRIBE لتقدير الوسط الحسابى بنقطة للعينة العشوائية البسيطة أو المنتظمة أو متوسط الطبقة :

```
DESCRIBE Variable
```

وسيعطى هذا الأمر المقاييس التالية :

n حجم العينة ويمثل عدد القيم التى تم إدخالها .
MEAN الوسط الحسابى .
MEDIAN الوسيط .
TRMEAN الوسط الحسابى بعد حذف نسبة من القيم الدنيا ونسبة من القيم العليا (٥٪ مثلا من عدد القيم العليا و٥٪ من عدد القيم الدنيا ومتوسط الباقى الذى يمثل ٩٠٪ من عدد القيم هو المتوسط TRMEAN) .
STDEV الانحراف المعيارى (S) .
SEMEAN الخطأ المعيارى (الانحراف المعيارى للمتوسط) .
MAX MIN أكبر قيمة وأصغر قيمة .
Q_1, Q_3 الربيع الأول والربيع الثالث .

ب - تقدير الوسط الحسابى بفترة

١ - عندما يكون تباين المجتمع (σ^2) معلوماً ، نستخدم ZINTERVAL وندخل الانحراف المعيارى σ كما يلى :

```
SET C1
data
END
ZINTERVAL [k  Percen Confidence]
SIGMA = K , C1
```

وسيعطى النتائج التالية :

N , MEAN , STDEV , SEMEAN Confidence intervals

(K هى النسبة المئوية لفترة الثقة أى ... K= 90 or 95 or)

٢ - عندما يكون تباين المجتمع مجهولاً ، نستخدم الأمر TINTERVAL كما يلى

```
SET C2
data
END
TINTERVAL [K  Percen Confidence] , C2
```

وسيعطى هذا الأمر النتائج المنوه عنها سابقًا .

تطبيق (١١ – ٣)

باستخدام بيانات التطبيق (١١ –٢) استخرج :

١ – الوسط الحسابى والوسيط والانحراف المعيارى لسنوات الخبرة .

٢ – تقدير حدى الثقة بمستوى ثقة (٩٥٪) .

أ – إذا كان الانحراف المعيارى لسنوات الخبرة (١,٥٨٩) .

ب – إذا كان الانحراف المعيارى مجهولاً .

```
MTB > read c1 - c3
MTB > data > 3 1 1
MTB > data > 4 2 2
MTB > data > 5 1 2
MTB > data > 6 1 1
MTB > data > 7 2 2
MTB > data > 8 1 1
 ....
....
MTB > data > 4 2 1
MTB > data > end
MTB > name c1 'exp'
MTB > name c2 'dep'
MTB > name c3 'city'

MTB > print c1 - c2
```

ROW	exp	dep	city
1	3	1	1
2	4	2	2
3	5	1	2
4	6	1	1
5	3	2	2
6	8	1	1
7	4	2	1
8	4	1	1
9	5	2	2
10	3	2	2
11	6	2	1
12	7	2	1
13	4	2	1

MTB > desc c1 - c2

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
exp	13	4.769	4.000	4.636	1.589	0.441
dep	13	1.615	2.000	1.636	0.506	0.140

	MIN	MAX	Q1	Q3
exp	3.000	8.000	3.500	6.000
dep	1.000	2.000	1.000	2.000

MTB > tinterval c1 - c2

	N	MEAN	STDEV	SE MEAN	95.0 PERCENT C.I.
exp	13	4.769	1.589	0.441	(3.809, 5.730)
dep	13	1.615	0.506	0.140	(1.309, 1.921)

MTB > zinterval sigma = 1.589 c1

THE ASSUMED SIGMA = 1.59

	N	MEAN	STDEV	SE MENA	95.0 PERCENT C.I.
exp	13	4.769	1.589	0.441	(3.904 , 5.634)

١١ - ٤ استخدام نظام ساس (SAS) فى مجال العينات

يحتوى نظام ساس على أوامر نستطيع استخدامها فى مجال العينات ويمكننا تلخيصها فى نوعين من الأوامر :

١ - أوامر توليد الأرقام العشوائية .

٢ - أوامر تتعلق بعرض البيانات جدولياً وبيانياً وتقدير معلمات المجتمع وتحليل البيانات .

ويعد نظام ساس ذا إمكانيات كبيرة خاصة إذا كان حجم المجتمع كبيراً والبيانات متعددة .

وسنقوم باستعراض الأوامر المتعلقة بتوليد الأرقام العشوائية واختيار وحدات العينة وعرض بياناتها جدولياً وتقدير معالم المجتمع .

١١-٤-١ توليد الأرقام العشوائية . (Generating Random Numbers)

باستخدام أوامر الأرقام العشوائية ، نستطيع توليد أرقام عشوائية لمختلف التوزيعات حيث تستخدم هذه الأوامر دليل (Argument) لاختيار مايسمى قيمة البذرة الأولية (Initial Seed Value) التى تنشئ مجرى الأرقام العشوائية واتجاهها . ويكون هذا الدليل صفراً أو عدداً أكبر من الصفر أو عدداً أصغر من الصفر .

وإذا أردنا التحكم فى اتجاهات متعددة للأرقام العشوائية يتم استخدام التعليمة CALL .

ولابد لنا من الإشارة إلى أن توليد الأرقام العشوائية يكون لعدة توزيعات كتوزيع ذى الحدين وتوزيع كوشى وتوزيع جاما والتوزيع الطبيعى والتوزيع المنتظم والتوزيعات الأخرى * .

ويستخدم أمر توليد الأرقام العشوائية للتوزيع المنتظم RANUNI (SEED)

لتكوين الأرقام العشوائية التى تستخدم لتحديد أرقام الوحدات المختارة للعينة العشوائية البسيطة والعينة الطبقية والعينة المنتظمة وأيضاً اختيار العناقيد عشوائياً فى العينة العنقودية .

ويمكننا اختيار وحدات العينة باستخدام إحدى الطرق التالية :

١ - لاختيار وحدات عينة عشوائية بسيطة ، نحدد القيمة العليا التى يأخذها المتغير للعينة المختارة ولتكن (٠,٢٥) **.

ونستخدم الأمر التالى :

```
DATA RAN;
SET BIGDATA;
IF RANINI (o) ≤ = .25 THEN OUTPUT;
```

إن الأمر (RANUNI) يعيد الرقم الذى تم توليده من التوزيع المنتظم فى الفترة (o, 1) وذلك لأية قيمة بذرة عددية باستخدام المولد الضربى الأولى (2^{31} - 1) والمضروب 397204094 والبذرة يجب أن يكونا ثابتاً عددياً أقل من (2^{31} - 1) *** .

ويوضح التطبيق رقم (١١-٤) استخدام هذه الأوامر لتوليد الأرقام العشوائية .

---

* Sas : User's Guide, Basic, (PP 261 - 271) .

** Aronson M. & A. : SAS SYSTEM, "A Programmer`s Guide" , McGraw Hill, Inc, 1990 (P.312) .

*** SAS : USER`S Guide , Basic (Ibd) (P.269) .

٢ – تستخدم الأوامر التالية لتوليد أرقام عشوائية لأكثر من مجرى أو اتجاه

```
DATA A:
RETAIN SEEDI SEED2 1613218064:
Do I = 1 TO 5 ;
X1 = RANUNI (SEED 1);
X2 = RANUNI (SEED 2);
OUTPUT ;
END ;
PROC PRINT :
TITLE 'USING A RANDOM NUMBERS FUNCTON' ;
```

إن الأرقام المولدة ستكون فى (X1) و (X2) التى تختلف أرقامها .

ويوضح التطبيق (١١–٥) استخدام هذه الأوامر لتوليد الأرقام العشوائية للعينة العشوائية البسيطة .

وإذا أردنا استخدام تعليمة CALL نضع التعليمتين التاليتين :

```
CAIL RANUNI (SEED 4 , x4) ;
CAIL RANUNI (SEED 5 , x5) ;
```

وذلك عوضاً عن التعليمتين التاليتين فى البرامج أعلاه :

```
x1 = RANUNI (SEED 1) ;
x2 = RANUNI (SEED 2) ;
```

تطبيق (١١ – ٤)

البيانات التالية تمثل أسماء منسوبى إحدى الجهات والمدن التى يعملون بها وجنسياتهم وأعمارهم . المطلوب اختيار عينة عشوائية إذا كان أعلى قيمة يأخذها المتغير المنتظم هى ٠,٢٥ .

OBS	NAME	CITY	NAT	AGE
1	ALI	A	SAUDI	31
2	SAD	B	NSAUDI	32
3	AHMWD	A	NSAUDI	33
4	SAMIR	C	NSAUDI	34
5	FADEE	B	SAUDI	35
6	SAUD	A	SAUDI	30
7	SAMEE	C	NSAUDI	30
8	SALEAH	B	SAUDI	37
9	ATA	A	NSAUDI	40
10	AHMAD	C	NSAUDI	44
11	ALI	B	SAUDI	60
12	SAD	B	NSAUDI	55
13	AHMWD	C	SAUDI	44
14	SAMIR	B	SAUDI	33
15	FADEE	C	NSAUDI	31

OBS	NAME	CITY	NAT	AGE
16	SAUD	A	SAUDI	32
17	SAAD	B	SAUDI	33
18	SAEED	A	NSAUDI	22
19	SALEAH	C	SAUDI	21
20	ATA	C	NSAUDI	25
21	IBRAHIM	A	NSAUDI	26
22	ALI	A	SAUDI	27
23	ALI	B	SAUDI	52
24	AHMWD	B	NSAUDI	42
25	SAMIR	B	NSAUDI	32
26	FADEE	A	SAUDI	18
27	SAUD	C	NSAUDI	18
28	SAMEE	C	SAUDI	22
29	SALEAH	C	NSAUDI	32
30	ATA	B	SAUDI	42
31	IBRAHIM	C	SAUDI	52
32	ALI	C	SAUDI	56
33	SAD	C	NSAUDI	42
34	AHMWD	A	SAUDI	32
35	SAMIR	B	SAUDI	32
36	FADEE	B	NSAUDI	21
37	SAUD	B	SAUDI	22
38	SAMEE	A	SAUDI	23
39	SALEAH	A	NSAUDI	24
40	ATA	A	SAUDI	25
41	IBRAHIM	A	SAUDI	22
42	ALI	C	NSAUDI	22
43	SAD	C	SAUDI	52
44	AHMWD	A	SAUDI	52
45	SAMIR	B	NSAUDI	52
46	FADEE	B	SAUDI	27
47	SAUD	B	SAUDI	22
48	SAMEE	A	NSAUDI	22
49	SALEAH	A	SAUDI	22
50	ATA	A	SAUDI	19

الحل :

نكتب البرنامج التالي :

```
INPUT NAME $ CITY $ 9 NAT $ 11 - 17
AGE 18 - 19 ;
CARDS ;
.
.
.
;
```

```
DATA RAN ;
SET BIGDATA ;
IF RANUNI (0) < = .25 THEN OUTPUT ;
PROC PRINT ;
```

وسنحصل على النتائج التالية :

OBS	NAME	CITY	NAT	AGE
1	SALEAH	B	SAUDI	37
2	SAUD	A	SAUDI	32
3	SAAD	B	SAUDI	33
4	ATA	B	SAUDI	42
5	ALI	C	SAUDI	56
6	IBRAHIM	A	SAUDI	22

تطبيق (١١ -٥)

باستخدام بيانات التطبيق (١١ - ٤) ماهى الأرقام المولدة عشوائيًا إذا كان المطلوب استخدام التوزيع المنتظم باتجاهين (مجريين) .

الحل :

نكتب البرنامج التالي :

```
DATA BIGDATA
INPUT NAMES CITYS 9 NATS 11 - 17
AGE 18 - 19 ;
CARDS ,
.
.
.
;
DATA RAN ;
SET BIGDATA ;
RETAIN SEED 1 SEED 2 1613218069 ;
DO I = 1 to 2 ;
X 1 = RANUNI (SEED1) ;
X 2 = RANUNI (SEED2) ;
OUTPUT ;
END ;
PROC PRINT ;
```

وسنحصل على النتائج التالية :



THE SAS SYSTEM

13 : 18 WEDNESDAY , JUNE 29 , 1994 (1)

OBS	NAME	CITY	NAT	AGE	SEED 1	SEED 2	I	X 1	X 2
1	ALI	A	SAUDI	31	1613218064	1613218064	1	0.80083	0.77094
2	ALI	A	SAUDI	31	1613218064	1613218064	2	0.00960	0.49815
3	SAD	B	NSAUDI	32	1613218064	1613218064	1	0.44219	0.64603
4	SAD	B	NSAUDI	32	1613218064	1613218064	2	0.50046	0.73160
5	AHMWD	A	NSAUDI	33	1613218064	1613218064	1	0.55804	0.50067
6	AHMWD	A	NSAUDI	33	1613218064	1613218064	2	0.22734	0.43086
7	SAMIR	C	NSAUDI	34	1613218064	1613218064	1	0.98315	0.90948
8	SAMIR	C	NSAUDI	34	1613218064	1613218064	2	0.77066	0.47913
9	FADEE	B	SAUDI	35	1613218064	1613218064	1	0.58164	0.50294
10	FADEE	B	SAUDI	35	1613218064	1613218064	2	0.32804	0.55016
11	SAUD	A	SAUDI	30	1613218064	1613218064	1	0.22125	0.31789
12	SAUD	A	SAUDI	30	1613218064	1613218064	2	0.33826	0.64132
13	SAMEE	C	NSAUDI	30	1613218064	1613218064	1	0.55537	0.11451
14	SAMEE	C	NSAUDI	30	1613218064	1613218064	2	0.92837	0.79580
15	SALEAH	B	SAUDI	37	1613218064	1613218064	1	0.14996	0.51984
16	SALEAH	B	SAUDI	37	1613218064	1613218064	2	0.56242	0.99012
17	ATA	A	NSAUDI	40	1613218064	1613218064	1	0.29886	0.92789
18	ATA	A	NSAUDI	40	1613218064	1613218064	2	0.78116	0.72669
19	AHMAD	C	NSAUDI	44	1613218064	1613218064	1	0.34420	0.06357
20	AHMAD	C	NSAUDI	44	1613218064	1613218064	2	0.96818	0.16312
21	ALI	B	SAUDI	60	1613218064	1613218064	1	0.49378	0.85501
22	ALI	B	SAUDI	60	1613218064	1613218064	2	0.71172	0.30749
23	SAD	B	NSAUDI	55	1613218064	1613218064	1	0.54433	0.92145
24	SAD	B	NSAUDI	55	1613218064	1613218064	2	0.22682	0.49549
25	AHMWD	C	SAUDI	44	1613218064	1613218064	1	0.18077	0.30401
26	AHMWD	C	SAUDI	44	1613218064	1613218064	2	0.95084	0.38794
27	SAMIR	B	SAUDI	33	1613218064	1613218064	1	0.20029	0.98540
28	SAMIR	B	SAUDI	33	1613218064	1613218064	2	0.07752	0.78539
29	SADEE	C	NSAUDI	31	1613218064	1613218064	1	0.31496	0.94053
30	SADEE	C	NSAUDI	31	1613218064	1613218064	2	0.37987	0.38912
31	SAUD	A	SAUDI	32	1613218064	1613218064	1	0.02803	0.26490
32	SAUD	A	SAUDI	32	1613218064	1613218064	2	0.81175	0.27905
33	SAAD	B	SAUDI	33	1613218064	1613218064	1	0.84944	0.57182
34	SAAD	B	SAUDI	33	1613218064	1613218064	2	0.52137	0.68865
35	A	A	NSAUDI	22	1613218064	1613218064	1	0.94549	0.77711
36	A	A	NSAUDI	22	1613218064	1613218064	2	0.42546	0.61510
37	SALEAH	C	SAUDI	21	1613218064	1613218064	1	0.57266	0.84416
38	SALEAH	C	SAUDI	21	1613218064	1613218064	2	0.54355	0.74819
39	ATA	C	NSAUDI	25	1613218064	1613218064	1	0.65674	0.98911
40	ATA	C	NSAUDI	25	1613218064	1613218064	2	0.24333	0.45421
41	IBRAHIM	A	NSAUDI	26	1613218064	1613218064	1	0.60199	0.00328
42	IBRAHIM	A	NSAUDI	26	1613218064	1613218064	2	0.02233	0.48033
43	ALI	A	SAUDI	27	1613218064	1613218064	1	0.14610	0.63396
44	ALI	A	SAUDI	27	1613218064	1613218064	2	0.13759	0.21971

OBS	NAME	CITY	NAT	AGE	SEED 1	SEED 2	I	X 1	X 2
45	ALI	B	SAUDI	52	1613218064	1613218064	1	0.64001	0.20404
46	ALI	B	SAUDI	52	1613218064	1613218064	2	0.55536	0.89610
47	AHMWD	B	NSAUDI	42	1613218064	1613218064	1	0.43239	0.11103
48	AHMWD	B	NSAUDI	42	1613218064	1613218064	2	0.90621	0.31173
49	SAMIR	B	NSAUDI	32	1613218064	1613218064	1	0.22358	0.15252
50	SAMIR	B	NSAUDI	32	1613218064	1613218064	2	0.14569	0.98341
51	FADLE	A	SAUDI	18	1613218064	1613218064	1	0.16486	0.41620
52	FADEL	A	SAUDI	18	1613218064	1613218064	2	0.57815	0.18747
53	SAUD	C	NSAUDI	18	1613218064	1613218064	1	0.28312	0.93688
54	SAUD	C	NSAUDI	18	1613218064	1613218064	2	0.13122	0.20668
55	SAMEE	C	SAUDI	22	1613218064	1613218064	1	0.45975	0.63771
56	SAMEL	C	SAUDI	22	1613218064	1613218064	2	0.81634	0.47135
57	SALEAH	C	NSAUDI	32	1613218064	1613218064	1	0.92210	0.31022
58	SALEAH	C	NSAUDI	32	1613218064	1613218064	2	0.50801	0.26337
59	ATA	B	SAUDI	42	1613218064	1613218064	1	0.85622	0.74093
60	ATA	B	SAUDI	42	1613218064	1613218064	2	0.34899	0.71644
61	IBRAHIM	C	SAUDI	52	1613218064	1613218064	1	0.75282	0.85185
62	IBRAHIM	C	SAUDI	52	1613218064	1613218064	2	0.67672	0.49787
63	ALI	C	SAUDI	56	1613218064	1613218064	1	0.20730	0.43161
64	ALI	C	SAUDI	56	1613218064	1613218064	2	0.95372	0.28160
65	SAD	C	NSAUDI	42	1613218064	1613218064	1	0.82440	0.02757
66	SAD	C	NSAUDI	42	1613218064	1613218064	2	0.45317	0.21084
67	AHMWD	A	SAUDI	32	1613218064	1613218064	1	0.13744	0.48517
68	AHMWD	A	SAUDI	32	1613218064	1613218064	2	0.01076	0.42092
69	SAMIR	B	SAUDI	32	1613218064	1613218064	1	0.93025	0.92262
70	SAMIR	B	SAUDI	32	1613218064	1613218064	2	0.42329	0.51621
71	FADEL	B	NSAUDI	21	1613218064	1613218064	1	0.99985	0.81537
72	FADEL	B	NSAUDI	21	1613218064	1613218064	2	0.80567	0.69397
73	SAUD	B	SAUDI	22	1613218064	1613218064	1	0.53509	0.02336
74	SAUD	B	SAUDI	22	1613218064	1613218064	2	0.85127	0.49641
75	SAMEE	A	SAUDI	23	1613218064	1613218064	1	0.25196	0.02110
76	SAMEE	A	SAUDI	23	1613218064	1613218064	2	0.23050	0.340 5
77	SALEAH	A	NSAUDI	24	1613218064	1613218064	1	0.33474	0.94498
78	SALEAH	A	NSAUDI	24	1613218064	1613218064	2	0.91298	0.03596
79	ATA	A	SAUDI	25	1613218064	1613218064	1	0.58587	0.96149
80	ATA	A	SAUDI	25	1613218064	1613218064	2	0.47680	0.30561
81	IBRAHIM	A	SAUDI	22	1613218064	1613218064	1	0.71805	0.02835
82	IBRAHIM	A	SAUDI	22	1613218064	1613218064	2	0.46764	0.86988
83	ALI	C	NSAUDI	22	1613218064	1613218064	1	0.19204	0.27279
84	ALI	C	NSAUDI	22	1613218064	1613218064	2	0.16287	0.85018
85	SAD	C	SAUDI	52	1613218064	1613218064	1	0.68168	0.61348
86	SAD	C	SAUDI	52	1613218064	1613218064	2	0.56461	0.27641
87	AHMWD	A	SAUDI	52	1613218064	1613218064	1	0.18455	0.50463

THE SAS SYSTEM

13 : 18 WEDNESDAY , JUNE 29 , 1994 (3)

OBS	NAME	CITY	NAT	AGE	SEED 1	SEED 2	I	X 1	X 2
88	AHMWD	A	SAUDI	52	1613218064	1613218064	2	0.91259	0.55425
89	SAMIR	B	NSAUDI	52	1613218064	1613218064	1	0.36463	0.49125
90	SAMIR	B	NSAUDI	52	1613218064	1613218064	2	0.06467	0.13484
91	FADEL	B	SAUDI	27	1613218064	1613218064	1	0.26409	0.04457
92	FADEL	B	SAUDI	27	1613218064	1613218064	2	0.47172	0.16686
93	SAUD	B	SAUDI	22	1613218064	1613218064	1	0.98951	0.77734
94	SAUD	B	SAUDI	22	1613218064	1613218064	2	0.93342	0.508
95	SAMEE	A	NSAUDI	22	1613218064	1613218064	1	.60021	0.028
96	SAMEE	A	NSAUDI	22	1613218064	1613218064	2	0.34580	0.53162
97	SALEAH	A	SAUDI	22	1613218064	1613218064	1	0.29402	0.14076
98	SALEAH	A	SAUDI	22	1613218064	1613218064	2	0.19037	0.83062
99	ATA	A	SAUDI	19	1613218064	1613218064	1	0.46912	0.16075
100	ATA	A	SAUDI	19	1613218064	1613218064	2	0.68014	0.71779

١١-٤-٢ طريقة اختيار العينة المنتظمة :

نستطيع اختيار وحدة من كل (k) سجل من جميع الوحدات . وعيب هذه الطريقة ضرورة قراءة كامل أوراق الملف واحدة بعد أخرى . إذا كنا مثلا نريد اختيار واحد من كل (٥) سجل (K = 5) .

```
DATA NTH;
   SET BIGDATA;
   RETAIN I 1;
   IF I = 5
      THEN DO;
      OUTPUT;
      I = 1;
      END ;
   ELSE I + 1;
```

ويوضح التطبيق رقم (١١ - ٦) استخدام هذه الطريقة .

١١-٤-٢ طريقة اختيار العينة الطبقية .

نستخدم الطريقة السابقة الموضحة لاختيار العينة العشوائية البسيطة ، وذلك على اعتبار أن كل طبقة من المجتمع تمثل مجتمعاً فرعياً نريد اختيار عينة منه .

وتبدأ عملية الاختيار الطبقي بتقسيم المجتمع ومن ثم اختيار عينة جزئية من كل طبقة باستخدام إحدى الطرق السابقة .

ويمكننا تقسيم وحدات المعاينة فى المجتمع إلى طبقات (STRATA) باستخدام أحد المتغيرات الرئيسية الذى يستخدم كمعيار للتقسيم الطبقى ، وقد يكون هذا المعيار اسمياً مثل المناطق الجغرافية أو الجنس أو الإدارة أو الوزارة ، وقد يكون هذا المعيار متغيراً كمياً مثل العمر أو الدخل أو غيرهما . إن البيانات والمعلومات المتعلقة بوحدات المعاينة التى يتضمنها الإطار ، غالباً ما تكون محفوظة فى ملف (أو وحدة البيانات) ، وللقيام بتصنيف الوحدات فى طبقات نستخدم إحدى الطريقتين التاليتين حسب المعيار المستخدم :

أ – إذا كان المعيار المستخدم اسمياً نستخدم الأمر SORT كما يلى :

```
PROC SORT ;
VAR Variable
```

مثلاً إذا كان المعيار المستخدم المدينة City يصبح الأمر :

```
PROC SORT ;
VAR CITY
```

ب – إذا كان المعيار المستخدم كمياً كالعمر أو الدخل أو غيرهما نستخدم

```
PROC LIFETEST ;
STRAT Variable   < interval List > ... ;
```

مثلاً إذا كنا نريد تصنيف البيانات حسب العمر والجنس نستخدم

```
PROC LIFETEST ;
STRAT AGE (15 to 65 BY 10) SEX ;
```

أى أن طول الفئة (١٠) سنوات تبدأ من العمر(١٥) سنة ، أو نستخدم الطريقة المستخدمة لعرض البيانات فى الفئات التى سنشرحها فى الصفحات القادمة .

تطبيق (١١ – ٦)

باستخدام بيانات التطبيق (١١ – ٤) ، نريد اختيار عينة منتظمة حجمها (١٠) موظفين (أى واحد من خمسة) .

الحل

نكتب البرنامج التالى :

```
DATA BIGDATA ;
INPUT NAMES CITYS 9 NATS 11 - 17 AGE 18 - 19 ;
CARDS ;
.
.
.
DATA NTH ;
SET BIGDATA ;
RETAIN I 1;
IF I = 5 THEN DO ;
OUTPUT ,
I = 1 ;
END ;
ELSE I + 1 ;
PROC PRINT ;
```

وسنحصل على النتائج التالية :

OBS	NAME	CITY	NAT	AGE	I
1	FADEE	B	SAUDI	35	5
2	AHMAD	C	NSAUDI	44	5
3	FADEE	C	NSAUDI	31	5
4	ATA	C	NSAUDI	25	5
5	SAMIR	B	NSAUDI	32	5
6	ATA	B	SAUDI	42	5
7	SAMIR	B	SAUDI	32	5
8	ATA	A	SAUDI	25	5
9	SAMIR	B	NSAUDI	52	5
10	ATA	A	SAUDI	19	5

١١-٤-٥ طريقة اختيار عينة غير عشوائية :

نستخدم الأوامر التالية لاختيار عدد صغير ومحدد من الوحدات يستخدم أحيانًا لاختيار خطوات تصميم البحث والاستمارة * :

```
DATA ;
SET JOBCODES (FIRSTOBS = 21 OBS = 50) ;
```

حيث سيقرأ الحاسوب (٢١) سجلاً من (٥٠) سجلاً مخزنة فى وحدة البيانات .

١١-٤-٦ عرض بيانات العينة جدوليا باستخدام (SAS)

بعد جمع البيانات من وحدات المعاينة المختارة ، يتم إدخال البيانات فى الحاسوب وذلك لعرضها جدوليًا ، ولاستخراج البيانات فى شكل جداول نستخدم الأمر .

```
PROC FREQ ;
TABLES Variabe ;
```

إذا كان الجدول بسيطًا أى يتعلق بظاهرة واحدة – أما إذا كان الجدول مركبًا فنستخدم الأمر :

```
PROC FREQ ;
TABLES Variable * Variable ;
```

تطبيق (١١ – ٧)

باستخدام بيانات التطبيق (١١ – ٤) ، المطلوب توزيع الموظفين حسب المدينة ، وتوزيعهم حسب الجنسية والمدينة ، وتوزيعهم حسب فئات العمر .

* ARONSON M&A : SAS SYSTEM A PROGRAMMER'S Guide (Ibd) (P.311) .

الحل

نستخدم الأوامر التالية للحصول على مايلي :

أ - توزيع الموظفين حسب المدن .

ب - توزيع الموظفين حسب الجنسية والمدن .

جـ - توزيع الموظفين حسب فئات العمر (خمس فئات تبدأ من الفئة (١٥-٢٤) وطول الفئة (١٠) سنوات .

```
PROC FORMAT ;
        VALUE AGROUP
                  15 - 24    = '15 - 24'
                  25 - 34    = '25 - 34'
                  35 - 44    = '35 - 44'
                  45 - 54    = '45 - 54'
                  55 - HIGH = '55 and over' ;
DATA FAHAD ;
INPUT NAMES CITY , NATS 11 - 17 AGE 18 - 19 ;
FORMAT AGE AGROUP. ;
CARDS ;
(data) .. ...
;
PROC SORT ;
BY CITY ;
BROC FREQ ;
                  TABLES CITY ,
                  TABLES NAT , * CITY
                  TABLES AGE ;
PROC PRINT ;
```

وسنحصل على النتائج التالية :

THE SAS SYSTEM

10 : 31 Sunday , June 21 , 1994

CITY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	6	30.0	6	30.0
B	7	35.0	13	65.0
C	7	35.0	20	100.0

TABLE OF NAT BY CITY

NAT Frequency Percent Row Pct Col Pct	CITY A	B	C	Total
NSAUDI	3 15.00 30.00 50.00	2 10.00 20.00 28.57	5 25.00 50.00 71.43	10 50.00
SAUDI	3 15.00 30.00 50.00	5 25.00 50.00 71.43	2 10.00 20.00 28.57	10 50.00
Total	6 30.00	7 35.00	7 35.00	20 100.00

AGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15 - 24	2	10.0	2	10.0
25 - 34	11	55.0	13	65.0
35 - 44	5	25.0	18	90.0
55 - AND OVER	2	10.0	20	100.0

OBS	NAME	CITY	NAT	AGE
1	ALI	A	SAUDI	25 - 34
2	AHMWD	A	NSAUDI	25 - 34
3	SAUD	A	SAUDI	25 - 34
4	ATA	A	NSAUDI	35 - 44
5	SAUD	A	SAUDI	25 - 34
6	SAEED	A	NSAUDI	15 - 24
7	SAD	B	NSAUDI	25 - 34
8	FADEE	B	SAUDI	35 - 44

OBS	NAME	CITY	NAT	AGE
9	SALLEAH	B	SAUDI	35 - 44
10	ALI	B	SAUDI	55 - AND OVER
11	SAD	B	NSAUDI	55 - AND OVER
12	SAMIR	B	SAUDI	25 - 34
13	SAD	B	SAUDI	25 - 34
14	SAMIR	C	NSAUDI	25 - 34
15	SAMEE	C	NSAUDI	25 - 34
16	AHMAD	C	NSAUDI	35 - 44
17	AHMAD	C	SAUDI	35 - 44
18	FADEE	C	NSAUDI	25 - 34
19	SALEAH	C	SAUDI	15 - 24
20	ATA	C	NSAUDI	25 - 34

١١-٤-٧ تقدير معلمات المجتمع

يستخدم الأمر MEANS لتقدير متوسط المجتمع بنقطة أو بفترة ثقة بمستوى ثقة محدد .
ويمكننا استخدام هذا الأمر لتقدير الوسط الحسابى للمجتمع من بيانات العينة العشوائية البسيطة أو العينة المنتظمة .

كذلك نستطيع تقدير هذا المتوسط للعينة الطبقية باستخراج متوسط الطبقة وتباينها وكتابة الأوامر التى تمكننا من تقدير المتوسط باستخدام الصيغ المناسبة .

أ تقدير معالم المجتمع باستخدام العينة العشوائية البسيطة :

إن الأمر المستخدم لتقدير متوسط المجتمع بنقطة من بيانات العينة العشوائية البسيطة :

```
PROC MEANS ;
VAR Variables ;
```

مثلاً لتقدير متوسط الراتب الشهرى ومتوسط العمر نستخدم الأمر :

```
PROC MEANS ;
VAR SALARY AGE ;
```

وستحصل على المتوسطات لكل من الراتب والعمر (MEANS) والانحراف المعيارى (STDDEV) والقيمة العليا للبيانات والقيمة الدنيا(MAXIMUM, MINIMUM)

أما إذا أردنا استخراج حدود الثقة بمستوى ثقة محدد $(1 - \alpha)\%$ فنستخدم الأمر :

```
PROC MEANS MEAN STDERR STDDEV VAR TLCLM UCLM ; VAR Variables ;
```

وسنحصل بهذا الأمر على متوسطات المتغيرات (MEANS)

وانحرافاتها المعيارية (STD) والخطأ المعيارى (STDERR) وحدى الثقة الأدنى والأعلى (LCLM , UCLM) والتباين (VAR) وقيمة إحصائية اختبار استيودنت (T TEST) وعدد القيم (N) .

تطبيق (١١ -٨)

سحبت عينة عشوائية حجمها (٣٠) موظفًا لتقدير متوسط العمر للموظف باستخدام طريقة التقدير بنقطة ، وطريقة التقدير بفترة (بمستوى ثقة ٩٥٪) . وكانت الأعمار كما يلي :
٣١ ، ٣٣ ، ٣٠ ، ٤٠ ، ٣٢ ، ٢٢ ، ٢٦ ، ٢٧ ، ١٨ ، ٣٢ ، ٢٥ ، ٣٧ ، ٦٠ ، ٥٥ ، ٣٣ ، ٣٣ ،
٥٢ ، ٤٢ ، ٣٢ ، ٤٢ ، ٣٠ ، ٤٤ ، ٤٤ ، ٣١ ، ٢١ ، ٢٥ ، ١٨ ، ٢٢ ، ٣٢ ، ٣٤ .

الحــل

نضيف إلى البرنامج الذى تم إعداده فى التطبيق (١١ – ٧) الأوامر التالية .

```
PROC MEANS ;
VAR AGE ;
PROC MEANS MEAN STDERR STD VAR
T LCLM UCLM MAXDEC = 3 ;
VAR AGE ;
PROC PRINT ;
```

وسنحصل على النتائج فى الصفحة التالية التى ستكون فى ثلاث منازل عشرية فقط .

ب - تقدير معالم المجتمع باستخدام العينة الطبقية

يتطلب تقدير معالم المجتمع من بيانات عينة طبقية استخراج متوسط كل طبقة والتباين والخطأ المعيارى باستخدام التعليمات نفسها التى استخدمناها لبيانات العينة العشوائية البسيطة (التطبيق ١١-٨) مع إدخال التعليمة المتعلقة بمعالجة بيانات الطبقة (PROC SORT) كما يتضح من التطبيق (١١ -٩) . وبعد استخراج التقديرات لكل طبقة نستخدم الصيغ المتعلقة بتقدير متوسط المجتمع وحدود الثقة من بيانات عينة طبقية .

THE SAS SYSTEM

10 : 32 Sunday , June 12 , 1994

Analysis Variable : AGE

N	Mean	Std Dev	Minimum	Maximum
30	33.7666667	10.2475677	18.0000000	60.0000000

THE SAS SYSTEM

10 : 32 Sunday , June 12 , 1994

Analysis Variable : AGE

Mean	N	Std Dev	Std Error	Variance	T
33.767	30	10.248	1.871	105.013	18.048

Lower 95.0% CLM	Upper. 95.0% CLM
29.940	37.593

OBS	NAME	CITY	NAT	AGE
1	ALI	A	SAUDI	31
2	AHMWD	A	NSAUDI	33
3	SAUD	A	SAUDI	30
4	ATA	A	NSAUDI	40
5	SAUD	A	SAUDI	32
6	SAEED	A	NSAUDI	22
7	IBRAHIM	A	NSAUDI	26
8	ALI	A	SAUDI	27
9	FADEE	A	SAUDI	18
10	SAD	B	NSAUDI	32
11	FADEE	B	SAUDI	35
12	SALEAH	B	SAUDI	37
13	ALI	B	SAUDI	60
14	SAD	B	NSAUDI	55
15	SAMIR	B	SAUDI	33
16	SAAD	B	SAUDI	33
17	ALI	B	SAUDI	52
18	AHMWD	B	NSAUDI	42
19	SAMIR	B	NSAUDI	32
20	ATA	B	SAUDI	42

OBS	NAME	CITY	NAT	AGE
21	SAMIR	C	NSAUDI	34
22	SAMEE	C	NSAUDI	30
23	AHMAD	C	NSAUDI	44
24	AHMWD	C	SAUDI	44
25	FADEE	C	NSAUDI	31
26	SALEAH	C	SAUDI	21
27	ATA	C	NSAUDI	25
28	SAUD	C	NSAUDI	18
29	SAMEE	C	SAUID	22
30	SALEAH	C	NSAUDI	32

تطبيق (١١ -٩)

باستخدام البيانات التالية التي سحبت من (٣) مدن لتقدير متوسط العمر بنقطة وبفترة ثقة باحتمال (٩٥٪) :

AGE	31	32	33	34	35	30	30	37	40	44
CITY	A	B	A	C	B	A	C	B	A	C
AGE	60	55	44	33	31	32	33	22	21	25
CITY	B	B	C	B	C	A	B	A	C	C

الحل :

نستخدم التعليمات التالية :

```
PROC SORT ;
BY CITY ;
PROC MEANS ;
CLASS CITY ;
VAR AGE ;
PROC MEANS MEAN STDERR STD VAR T LCLM UCLM
MAXDEC = 3 ;
CLASS CITY ;
VAR AGE ;
```

وسنحصل على النتائج التالية :

THE SAS SYSTEM

10 : 32 Sunday , June 12 , 1994

Analysis Variable : AGE

CITY	N Obs	N	Mean	Std Dev	Minimum	Maximum
A	6	6	31.3333333	5.7850382	22.0000000	40.0000000
B	7	7	40.7142857	11.6721076	32.0000000	60.0000000
C	7	7	32.7142587	8.7885207	21.0000000	44.0000000

THE SAS SYSTEM

10 : 32 Sunday , June 12 , 1994

Analysis Variable : AGE

CITY	N Obs	Mean	N	Std Dev	Std Error	Variance
A	6	31.333	6	4.785	2.362	33.467
B	7	40.714	7	11.672	4.442	136.238
C	7	32.714	7	8.789	3.322	77.238

CITY	N Obs	T	Lower 95.0% CLM	Upper 95.0% CLM
A	6	13.267	25.262	37.404
B	7	9.229	29.919	51.509
C	7	9.849	24.586	40.842

OBS	NAME	CITY	NAT	AGE
1	ALI	A	SAUDI	31
2	SAD	B	NSAUDI	32
3	AHMWD	A	NSAUDI	33
4	SAMIR	C	NSAUDI	34
5	FADEE	B	SAUDI	35
6	SAUD	A	SAUDI	30
7	SAMEE	C	NSAUDI	30
8	SALEAH	B	SAUDI	37
9	ATA	A	NSAUDI	40
10	AHMAD	C	NSAUDI	44
11	ALI	B	SAUDI	60
12	SAD	B	NSAUDI	55
13	AHMWD	C	SAUDI	44
14	SAMIR	B	SAUDI	33
15	FADEE	C	NSAUDI	31
16	SAUD	A	SAUDI	32
17	SAAD	B	SAUDI	33
18	SAEED	A	NSAUDI	22
19	SALEAH	C	SAUDI	21
20	ATA	C	NSAUDI	25

جـ - تقدير معالم المجتمع باستخدام العينة المنتظمة :

نستخدم التعليمات نفسها المستخدمة في تقدير معالم المجتمع لبيانات العينة العشوائية البسيطة (الفقرة أ) ولاستخراج حدى الثقة نقوم باستخدام الصيغة المناسبة بعد الحصول على التقديرات .

د - تقدير معالم المجتمع باستخدام العينة العنقودية :

يفضل تقدير متوسط العناقيد النهائية باستخدام التعليمات المتعلقة بكل عنقود ، ومن ثم نقوم باستخدام الصيغ المناسبة لتقدير متوسط المجتمع من بيانات العينة العنقودية .

أخيراً ، لابد لنا من الإشارة إلى أن استخدام برامج ساس ، يتطلب خبرة ومعارف بالأساليب الإحصائية والحاسوب ، وذلك لتنفيذ البرامج بالسرعة والدقة المناسبين .

الفصل
الثانى عشر

حالة عملية
عن استخدام العينات فى مجال البحوث



استعرضنا في الفصول السابقة أنواع العينات وكيفية اختيار وحداتها وتقدير معالم المجتمع من بيانات العينة باستخدام الحاسوب مع توضيح ذلك بتطبيقات متعددة .

وسنعالج في هذا الفصل حالة عملية شاملة تتعلق بتنفيذ بحث إحصائي بأسلوب المعاينة ، وذلك لإعطاء القارئ صورة شاملة عن مراحل البحث والخطوات التي تتكون منها .

لنفترض أنه طلب منك إجراء بحث اجتماعي يتعلق بانتشار ظاهرة تعاطي المخدرات ، وذلك لتقدير عدد متعاطي المخدرات وتوزيعاتهم المختلفة للتعرف على خصائصهم وتحديد الأسباب التي دفعتهم لتعاطي المخدرات واقتراح الحلول التي تقضي على هذه الآفة الخطيرة .

قبل اتخاذ قرار نهائي بإجراء بحث ميداني لجمع البيانات المطلوبة ، لابد من اتباع بعض الإجراءات وذلك لتقرير ما إذا كان تنفيذ هذا البحث يتطلب جمع البيانات ميدانيًا .

تتلخص الإجراءات الواجب اتباعها قبل تنفيذ البحث فيما يلي :

١ - تحديد مصادر البيانات :

قبل البدء في إجراء البحث ، يجب الاستعانة بما يلي :

- النشرات والتقاريرالسنوية الصادرة عن وزارة الداخلية ومصلحة الإحصاءات العامة ووزارة العدل للتعرف على أعداد المحكومين بجريمة تعاطي المخدرات خلال السنوات السابقة وأهم المعلومات والبيانات المتعلقة بهم .
- السجلات المتوافرة لدى بعض الجهات والتي تتضمن بيانات عن جرائم متعاطي المخدرات وليست منشورة .
- الدراسات السابقة التي عالجت موضوع الدراسة (إذا كانت متوافرة) والإجابة عن الأسئلة التالية :
  - هل هذه الدراسة أو الدراسات جديدة أم قديمة؟ .
  - هل محتويات هذه الدراسات ونتائجها مناسبة أم غير مناسبة ، وهل تتضمن جميع البيانات والمعلومات المطلوبة في الدراسة المزمع تنفيذها .
  - ما هي الإمكانات المالية والبشرية المتوافرة لإجراء بحث ميداني إذا تقرر ذلك .

على ضوء هذا يتم تقرير ما إذا كنا سنقوم بتنفيذ بحث ميداني لجمع البيانات ووضعها أو أنه لايوجد ضرورة لذلك لتوافر البيانات والمعلومات المطلوبة .

لنفترض أنه تقرر تنفيذ بحث ميداني وطلب منك القيام بذلك .

حينئذ يمكننا تقسيم مراحل تنفيذ البحث إلى خمس مراحل رئيسية :

١ – المرحلة التحضيرية أو ما يسمى مرحلة تصميم البحث .

٢ – مرحلة جمع البيانات ويتم فيها جمع البيانات من الوحدات الإحصائية .

٣ – المرحلة التجهيزية حيث يتم فيها إدخال البيانات على الحاسب وعرض البيانات جدوليًا وبيانيًا .

٤ – مرحلة وصف البيانات وتحليلها حيث يتم استخراج أهم المقاييس وإجراء الاختبارات المناسبة واستخدام الأساليب الإحصائية الأخرى لتحليل البيانات واستخلاص النتائج واقتراح التوصيات .

٥ – نشر البحث بالشكل المناسب .

وسنقوم بإجراء البحث مع ملاحظة أن البيانات والمعلومات الواردة فى الصفحات القادمة افتراضية .

١٢ – ١ مرحلة تصميم البحث :

مقدمة :

أصبحت المخدرات خطرًا يهدد الكثير من دول العالم بالدمار . وتنبهت الكثير من الدول إلى هذا الخطر فقامت ببذل الكثير من الجهود للحد من انتشار تعاطى المخدرات وتحديد أسباب انتشارها وإيجـــاد الحلول المناسبة .

وقد لوحظ فى السنوات الأخيرة ، انتشار ظاهرة تعاطى المخدرات وازدياد عدد متعاطيها فى عدد من الدول العربية ومن ضمنها المملكة ، وذلك على الرغم من القيود والإجراءات التى تتخذها الحكومة للحد من انتشار هذه الآفة الخطيرة . وقد قام بعض الباحثين بإجراء الدراسات لتحديد خصائص متعاطى المخدرات وتوزيعاتهم المختلفة وتحديد الأسباب التى شجعتهم على تعاطى المخدرات للعمل على إيجاد الحلول المناسبة .

١ – تحديد المشكلة :

يمكننا صياغة المشكلة بالسؤال التالى :

هل الخصائص الاجتماعية والمادية لمتعاطى المخدرات هى من الأسباب الرئيسية التى تؤدى إلى الوقوع فى هذه الآفة الخطيرة ؟ .

٢ - أهداف البحث :

يهدف البحث بشكل عام إلى التعرف على خصائص متعاطى المخدرات والأسباب التى دفعتهم لتعاطيها ، وإيجاد الحلول المناسبة للحد من هذه الظاهرة الخطرة .

وتتلخص الأهداف التفصيلية للبحث بما يلى :

- تحديد الخصائص الاجتماعية والمادية لمتعاطى المخدرات وتوزيعاتهم المختلفة حسب الحالة الزواجية والجنس والجنسية والعمر والحالة التعليمية والدخل ، ... .
- التعرف على الأسباب التى تشجع الأشخاص على تعاطى المخدرات .
- استخراج أهم المقاييس المتعلقة بخصائص متعاطى المخدرات
- اقتراح الحلول المناسبة للحد من انتشار تعاطى المخدرات .

٣ - شمول البحث :

يشمل البحث بعض السجناء المحكومين فى جريمة تعاطى المخدرات (عينة من السجناء) يتم اختيارهم من ثلاثة سجون تقع فى مدن الرياض وجدة والدمام سواء كانوا سعوديين أو غير سعوديين .

٤ - موعد تنفيذ البحث :

تقرر أن يكون موعد تنفيذ البحث هو ١/٤/........ وذلك لكون هذا التاريخ مناسباً سواء للمدلين بالبيانات أو الباحثين .

٥ - الوحدة الإحصائية (وحدة المعاينة)

وحدة المعاينة هى السجين الذى صدر بحقه حكم بجريمة تعاطى المخدرات سواء كان ذكراً أو أنثى وسواء كان سعودياً أو أجنبياً ، من نزلاء سجون مدن الرياض وجدة والدمام .

٦ - المجتمع الإحصائى :

هو جميع السجناء المحكوم عليهم بجرائم تعاطى المخدرات الذكور والإناث سواء كانوا سعوديين أم أجانب المسجونين فى سجون مدن الرياض وجدة والدمام .

٧ - الإطار الإحصائى :

تم إعداد ثلاث قوائم بأسماء جميع السجناء وأرقام ملفاتهم وغرفهم وعناوين عملهم وسكنهم ومدة الحكم الصادر بحقهم ، حيث تخصص قائمة لكل سجن من السجون الثلاثة فى

مدن الرياض وجـدة والدمام . ونورد فيما يلي إطار سجناء مدينة الرياض كمثال على هذه الإطارات .

إطار السجناء المتعاطين للمخدرات
في سجن مدينة .....

الرقم	الاسم	رقم الملف	رقم الغرفة	عنوان العمل	عنوان السكن	مدة الحكم
١	.....	٢١٢١	١٢	الرياض	الرياض	٣ سنوات
٢	....	١٢٢٢	١٢	الرياض	الرياض	٥ سنة
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

٨ - فروض البحث : تتلخص فروض البحث فيما يلي :

- غالبية متعاطى المخدرات هم من الأجانب الذين قدموا إلى المملكة .
- غالبية متعاطى المخدرات هم من الذكور .
- تعد العوامل المادية والعوامل الأسرية الاجتماعية من أهـم أسباب تعاطى المخدرات .

٩ - أسلوب جمع البيانات :

تبين من سجلات السجون في المدن الثلاث أن عدد السجناء المحكوم عليهم بجريمة تعاطى المخدرات كانت كما يلي بتاريخ ١/٣/........ (أرقام افتراضية) :

المدينة	عدد السجناء (N_h)	(النسبة المئوية $W_h = \frac{N_h}{N}$)
الرياض	٢١٠	٣٥٪
جدة	٢٤٠	٤٠٪
الدمام	١٥٠	٢٥٪
المجموع (N)	٦٠٠	١٠٠٪

وسيتبع أسلوب المعاينة كأسلوب لجمع البيانات حيث سيتم اختيارعينة طبقية عشوائية يتم توزيعها بطريقة التخصيص المتناسب على سجون المدن الثلاث .

١٠- طريقة جمع البيانات :

تقرر اتباع طريقة المراسلة كطريقة لجمع البيانات نظراً لحساسية الموضوع وعدم الرغبة فى إزعاج السجناء وإحراجهم بالأسئلة .

ولتوخى تعاون السجناء وضمان الإدلاء بإجاباتهم بدقة تقرر عدم طلب ذكر اسم السجين فى الاستبانة .

١١ - تحديد البيانات المطلوب جمعها :

لقد تم تحديد البيانات المطلوب جمعها على ضوء أهداف البحث وفروضه وطرق التحليل التى ستستخدم (التى تم تحديدها من قبل الباحث الذى سيقوم بالدراسة) وتتلخص هذه البيانات بما يلى :

- الجنسية
- العمر
- الحالة التعليمية
- نوع المخدرات
- الدخل
- مكان الإقامة الدائمة
- الجنس
- الحالة الزواجية
- المهنة
- أسباب تعاطى المخدرات
- عدد أفراد الأسرة

....

١٢ - الجداول :

نورد فيما يلى نماذج الجداول التى ستظهر فيها النتائج .

جدول رقم (١)

توزيع السجناء متعاطى المخدرات حسب الجنسية ومكان الإقامة بتاريخ / /

المدينة	سعودى		غير سعودى		المجموع	
	العدد	٪	العدد	٪	العدد	٪
الرياض جدة الدمام						
المجموع						

جدول رقم (٢)

توزيع السجناء متعاطى المخدرات حسب الجنس ومكان الإقامة بتاريخ / /

الجنس / مكان الإقامة	ذكور		إناث		الإجمالى	
	العدد	/	العدد	٪	العدد	٪
الرياض						
جدة						
الدمام						
المجموع						

جدول رقم (٣)

توزيع السجناء متعاطى المخدرات حسب الحالة الزواجية والجنس بتاريخ / /

الجنس / الحالة الزواجية	ذكور		إناث		الإجمالى	
	العدد	٪	العدد	٪	العدد	٪
متزوج						
لم يتزوج إطلاقًا						
أرمل						
مطلق						
المجموع						

(تابع إعداد نماذج الجدول)

....

....

....

١٢- تحديد حجم العينة وتوزيعها على السجون :

يتكون المجتمع من ثلاث طبقات ، ولتحديد حجم العينة تم اختيار عينة استطلاعية لاختبار الاستبانة وخطوات تصميم البحث واستنتاج بعض البيانات التي تساعد في تحديد حجم العينة ، وقد تم اختيار متوسط عمرالسجين في السجون الثلاثة لتحديد حجم العينة حيث تم اختيار (١٤) سجيناً (٥ سجناء من الرياض ، ٥ سجناء من جدة ، ٤ سجناء من الدمام) وتبين أن الأوساط الحسابية والانحرافات المعيارية للعمر كانت كما يلي :

	RIYD	JED	DAM
$\overline{X}_h$ (MEAN)	25	28	26
$\hat{\sigma}_h^2$ (Variance)	36	49	36
N_h (pop . size)	210	240	150
$W_h = \frac{N_h}{N}$	0.35	0.40	0.25

وبافتراض أن الخطأ المسموح به يساوى ($\beta = 2$) نستخدم الصيغة التالية :

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 \ \hat{\sigma}_h^2}{W_h}}{N^2 D + \sum_{h=1}^{L} N_h \ \hat{\sigma}_h^2}$$

إن

$$D = \frac{\beta^2}{4} = \frac{4}{4} = 1$$

البسط يساوى :

$$\frac{210^2 \times 36}{0.35} + \frac{240^2 \ \times 49}{0.40} \ \frac{150^2 \times 36}{0.25}$$

$= 4536000 + 7056000 + 3240000$

$= 14832000$

المقام يساوى :

$$(600^2 \times 1) + \{(210 \times 36) + (240 \times 49) + (150 \times 36)\}$$
$$= 360\,000 + 75560 + 11760 + 5400$$
$$= 384720$$

ويكون حجم العينة :

$$n = \frac{14832000}{384720} \simeq 39$$

أى أن حجم العينة (٣٩) سجيناً ويتم توزيعه على السجون الثلاثة باستخدام الصيغة التالية :

$$n_h = n\,\frac{N_h}{N} = nw_h$$
$$n_1 = 39 \times 0.35 = 14$$
$$n_2 = 39 \times 0.40 = 16$$
$$n_3 = 39 \times 0.25 = 9$$

وتم اختيار وحدات العينة باستخدام جداول الأرقام العشوائية ، وتبين أن أرقام السجناء المختارين كانت كما يلى :

الرياض : ١١٠ ، ١٢٠ ، ١١ ، ٤٨ ، ٥٦ ، ١٧٨ ، ٧ ، ٥٤ ، ٧٨ ، ١٩ ، ١٥٤ ، ٢١٨ ، ٨ ، ١٥ .

جدة : ٦٥ ، ٤٤ ، ٢٣٦ ، ٢٤ ، ١٣٦ ، ٢٣ ، ٢١٥ ، ٣٦ ، ٧٤ ، ٨٨ ، ١٤٨ ، ٢٢٠ ، ١٣٣ ، ١٦ ، ٩٩ ، ١٤٤ .

الدمام : ١٣٩ ، ٣٠ ، ٨ ، ١١٥ ، ٧٨ ، ٦٧ ، ١٣٨ ، ١٤ ، ١١٦ .

١٤ – الدعاية للبحث :

لقد تم إرسال خطاب مرفق به هدية إلى السجناء الذين تم اختيارهم كعينة بالبريد ، موضح فيه أهداف البحث وموعد إرسال الاستبانة وأهم المعلومات المتعلقة بالبحث ، وذلك لكسب ثقة السجناء للإدلاء بإجابات دقيقة .

١٥ - الخطة الزمنية :

لقد تم وضع خطة زمنية لتنفيذ البحث تتضمن مواعيد إجراء كل خطوة ومرحلة .

الخطة الزمنية لبحث أسباب تعاطى المخدرات

البيان	عدد الأيام	تاريخ البدء	تاريخ الانتهاء	ملاحظات
١ - تصميم البحث	١٥	/٣/١	/٣/١٥ .	
- تحديد المشكلة	١	/٣/١	/٣/١	
- تحديد الأهداف	١			
- تصميم الاستبانة	٤	/٣/١١	/٣/١٤	
٢ - جمع البيانات	٢٠	/٤/١	/٤/٢٠ . .	
- إرسال الاستبانات وإعادتها	١٦	/٤/١	٤/١٦	
- تدقيق الاستبانات	٤	٤/١٧	٤/٢١	
٣ - تجهيز البيانات	٥	٤/٢٥	٤/٣٠	
- إدخال البيانات على الحاسب	٥	٤/٢٥	٤/٣٠	
٤ - وصف وتحليل البيانات	٢٠	/٥/١	/٥/٢٠	
٥ - نشر النتائج	١٠	٦/١	٦/١٠	

١٦- ميزانية البحث :

لقد تم تخصيص (١٠٠٠٠) ريال لتغطية نفقات البحث الموضحـــة فى الميزانية التالية .

ميزانية بحث أسباب تعاطى المخدرات ...

المبالـــــغ				البيـــــان	تاريخ الإنفاق المتوقع
				رواتب وأجور	٣/١ - ٦/١
			×	......	
			×	......	
			×	......	
	xxxx			قرطاسية ومطبوعات	/٣/١٥
			×	أقلام	
	xxxx		×	طباعة استمارات	/٤/٢٥ ......
	xxx			نفقات حاسب آلي	/٦/١ ....
	xxx			طباعة ونشر النتائج	
	xxx				
٠٠	١٠٠٠٠			الإجمالى	

المملكة العربية السعودية

وزارة ....

البيانات الواردة فى الاستبانة سرية ولن تستخدم إلا للأغراض الإحصائية .

استبانة

بحث أسباب تعاطى المخدرات

فى مدن الرياض وجدة والدمام

ربيع الآخر ....

التاريخ : ٢٣/٤/ ٠٠٠٠٠٠٠

المكرم / ٠٠٠٠٠٠٠٠٠٠٠٠٠٠

بعد التحية ،،،

تقوم وزارة ٠٠٠٠٠٠٠ بإجراء بحث عن أسباب تعاطى المخدرات للحد منها ، وقد تم اختيارك عشوائيًا للإجابة على بعض الأسئلة للتعرف على هذه الأسباب والحد منها .

ويهدف البحث إلى :

- تحديد الخصائص الاجتماعية والمادية (الحالة الزواجية ، الجنس ، الجنسية ، العمر الحالة التعليمية ، الدخل) .
- تحديد الأسباب التى تشجع الأشخاص على تعاطى المخدرات .
- استخراج أهم المقاييس المتعلقة بخصائص متعاطى المخدرات ومقارنتها مع الأخرى .

يرجى قراءة جميع الأسئلة قبل الإجابة ، ومن ثم وضع إشارة (√) فى المربع الذى يتفق وإجابتك ، علمًا بأن البيانات الواردة فى هذه الاستبانة سرية ولن تستخدم إلا للأغراض الإحصائية لذا توخينا عدم ذكر الأسماء .

كما يرجى إعادة الاستبانة بعد وضعها بالظرف المرفق إلى العنوان المدون عليه فى موعد أقصاه ٥/٤/ ٠٠٠٠٠ ويرجى فى حالة الاستفسار الاتصال بالهاتف ٠٠٠٠٠٠٠ تحويلة ٠٠٠٠ .

شاكرين حسن تعاونكم ،،،

٠٠٠٠٠٠٠٠٠٠٠٠

المملكة العربية السعودية
وزارة ....

استبانة بحث
أسباب تعاطى المخدرات

القسم الأول : بيانات عامة :

١١ – العمـر [ ] سـنـة
٢١ – الجنس [١] ذكـر [٢] أنـثـى
٣١ – الجنسية [١] سـعـودى [٢] غير سعودى
٤١ – مكان الإقامة الدائمة : مدينة .... [ ]

القسم الثانى : الخصائص الاجتماعية والمادية :

١٢ – الحالة الزواجية :

[١] مـتـزوج [٢] غير متزوج
[٣] أرمـل [٤] مـطـلـق

٢٢ – المهنة .... [ ]

(تذكر المهنة · طالب أو تاجر أو عامل أو ....)

٢٣ – عدد أفراد الأسرة .... شخص (الذين يقيمون معك بشكل دائم)

٢٤ – الحالة التعليمية :

[١] أمـى [٢] يقرأ ويكتب
[٣] إبتـدائيـة [٤] مـتـوسطة
[٥] ثـانـويـة [٦] دبلوم فوق الثانوية
[٧] جـامـعى [٨] فوق الجامعى

٢٥ – الدخل الشهرى (يقصد بالدخل الشهرى الراتب الشهرى أو أى دخل آخر من عقارات أو تجارة أو أى مصدر آخر) .... ريالا / ريال .

القسم الثالث : أسباب تعاطى المخدرات :

١٢ - نوع المخدرات التى كنت تتعاطاها :

١	هيـــروين	٢	كـوكـائين
٣	أفـــيـون	٤	حـشـيش
٥	أخرى ، حدد		

٢٣ - حدد مما يلى الأسباب التى جعلتك تتعاطى المخدرات :

١ انخفاض مستوى الدخل .

٢ مشكلات اجتماعية تتعلق بالأسرة .

٣ عدم وجود عمل (البطالة) .

٤ وجود وقت فراغ كبير .

٥ أسباب أخرى حدد :

....

....

شكرًا على تعاونكم

١٢ - ٢ مرحلة جمع البيانات :

بتاريخ ١/٤/ .... وزعت الاستبانات بالبريد على السجناء فى المدن الثلاث واستغرقت عملية إعادتها بعد ملئها (١٥) يومًا وقد تم تدقيق الاستبانات بشكل سريع للتأكد من استلامها بشكل كامل وعدم وجود أسئلة لم يتم الإجابة عنها .

١٢ - ٣ مرحلة تجهيز البيانات :

تم تدقيق الإجابات للتأكد من ترابطها وعدم وجود تناقض فيها وتم إدخالها فى الحاسوب حيث تم تبويب البيانات باستخدام برامج ساس واستخرجت بيانات الجداول والرسوم البيانية وتم إعداد الجداول النهائية .

١٢ - ٤ مرحلة وصف وتحليل البيانات :

بعد إدخال البيانات بالحاسوب تم تقدير بعض المتوسطات والنسب وذلك كما يلى (كما هو موضح فى نهاية هذا الفصل) :

أ - تقدير متوسط عمر السجين المحكوم عليه . بجريمة تعاطى المخدرات :

	Riyd	Jedh	Damm	Total
$\overline{x}_h$ (Mean)	24	23	24.11	
$\hat{\sigma}^2_h$ (Variance)	42.308	35.600	37.611	
n_h (Sample Size)	14	16	9	39
N_h (pop . size)	210	240	150	600

ولتقدير متوسط عمر السجين الذى يتعاطى المخدرات نستخدم الصيغة التالية :

$$\overline{x} = \overline{x}_{prop} = \frac{\sum_{h=1}^{L} n_h \overline{x}_h}{n}$$

$$= \frac{\{24 \times 14\} + \{23 \times 16\} + \{24.11 \times 9\}}{39}$$

$$= \frac{921}{39} = 23.615$$

ب – تقدير حدى الثقة لمتوسط العمر .

$$\overline{x}_{st} \mp Z \sqrt{\hat{V}(\overline{x}_{st})}$$

$$\hat{V} \quad (\overline{x}_{prop}) = \frac{N-n}{N} \quad \sum_{h=1}^{L} \frac{N_h}{N} \frac{s_h^2}{N}$$

$$= \frac{600 - 39}{600} \left[\frac{\{(210 \times 42.308)\}}{600 \times 39} + \frac{\{240 \times 35.600\}}{600 \times 39} + \frac{\{150 \times 37.611\}}{600 \times 39} \right]$$

$$= \frac{0.935}{39} \{ 14.808 + 14.24 + 9.403 \} = 0.92$$

وبالتالى يكون حدا الثقة بمستوى ثقة (٩٥٪)

$23.615 \mp 1.96 \sqrt{0.92}$

23.615 ∓ 1.88

ويكون الحد الأدنى للعمر 21.74 سنــة

والحــد الأعلـى للعمــر 25.50 سنــة

أى أن متوسط المجتمع (متوسط عمر السجناء) سيقع بين هذين الحــدين بمستوى ثقة (٩٥٪) :

$21.74 \le \mu \le 25.5$

ج – لتقدير نسبة غير السعوديين المحكومين بجريمة تعاطى المخدرات بمستوى ثقة (٩٥٪) ، نستخدم الصيغة التالية لاستخراج حدى الثقة :

$$p_{st} \mp Z_{(1-\alpha/2)} \sqrt{\hat{V} \ (p_{st})}$$

$$p_{st} = \frac{\sum_{h=1}^{L} N_h \ p_h}{N}$$

من بيانات الحاسوب نجد أن :

$p_2 = \frac{5}{14} = 0.356 \qquad p_2 = 0.563 \qquad p_3 = 0.556$

$n_1 = 14 \qquad n_2 = 16 \qquad n_3 = 9$

$$p'_{st} = \frac{\{210 \times 0.357\} + \{240 \times 0.563\} + \{051 \times 0.556\}}{600}$$

$$= \frac{293.49}{600} = 0.489$$

$$\hat{V}(p_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h - 1} N_h^2 \frac{p_h q_h}{n_h}$$

$$= \frac{1}{\{600^2\}} \left[\frac{(210 - 14)^9}{210 - 1} \frac{\{210\}^2 \{0.356\} \quad \{0.564\}}{14} + \frac{240 - 16}{240 - 1} \{240\}^2 \frac{\{0.563\}\{0.437\}}{16} + \frac{150 - 9}{150 - 1} \{150\}^2 \frac{(0.556 \times 0.444)}{9} \right]$$

$$= \frac{1}{360\,000} \{678.11 + 830.12 + 584.02\}$$

$$= \frac{2092.25}{360\,000} = 0.00581$$

ويكون حدا الثقة بمستوى ثقة (0.95) .

$0.489 \mp 1.96 \sqrt{0.00581}$

0.489 ∓ 0.15

ويكون الحد الأدنى 0.339 أى ٣٣,٩٪

والحد الأعلى 0.639 أى ٦٣,٩٪

ويمكننا استخراج تقديرات متوسط الدخل والنسب الأخرى باستخدام الطريقتين السابقتين .

ثم نقوم باختبار فرضيات البحث واتخاذ القرارات المناسبة باستخدام الأساليب الإحصائية المناسبة والوصول إلى الاقتراحات .

١٢ - ٥ إعداد التقرير :

يتم إعداد التقرير النهائى الذى يتضمن خطة البحث والجداول وأهم المقاييس التى تم التوصل إليها والاختبارات والنتائج والتوصيات .

١٢ - ٦ يتم طباعة التقرير بشكل جيد وواضح .

نورد فيما يلى الأوامر التى استخدمت فى النتائج التى تم الوصول إليها باستخدام نظام (ساس) .

THE SAS SYSTEM

```
NOTE   COPYRIGHT (C) 1989 BY SAS INSTITUTE INC , CARY,NC USA
NOTE . SAS (R) PROPRIETARY SOFTWARE RELEASE 6.08 TS405
         LICENSED TO SAS INSTITUTE TRIAL SITE. SITE 0028350001

NOTE : RUNNING ON IBM  MODEL 5890 SERIAL NUMBER 020456

WELCOME TO THE SAS INFORMATION DELIVERY SYSTEM

GAS RELEASE 6.08
I. P. A                              SDD@P) : 0 S50380 3DDDDD ( )  @
INSTALLING DATE OF SAS R 6.8         8 FEB. 1994

NOTE : THE  SASUSER LIBRARY WAS NOT SPECIFIED . SASUSER LIBRARY  WILL
         NOW BE THE SAME
NOTE    ALL DATA SETS AND CATALOGS IN THE SASUSER LIBRARY WILL BE
         DELETED AT THE ENPREVENT THEIR DELETION .

NOTE    SAS SYSTEM OPTIONS SPECIFIED ARE
         SORT = 20 SORTWKON = 5

NOTE    THE INITIALIZATION PHASEUSED 0.14 CPU SECONDS AND 1942 K .
         OPTION LS = 80 ;

         00011018
         PROC FORMAT .

0020000
0020000
0030004
                    VALUE AGROUP 15 - 19 = ' 15 - 19 '
0030004
                                 20 - 24 = ' 20 - 24 '
0040004
                                 25 - 29 = ' 25 - 29 '
0050004
                                 30 - 34 = ' 30 - 34 '
0060004
                                 35 - HIGH = ' 35 AND OVER ' ;
0070004

ONTE : FORMAT AGROUP HAS BEEN OUTPUT .

0070004

NOTE · THE PROCEDURE FORMAT USED 0.03 CPU SECONDS AND 2054 K
         PROC FORMAT .

0072010
                    VALUE INCGRP 2000 - 2999 = ' 2000 - 2999 '
0072010
0                                3000 - 3999 = ' 3000 - 3999 '
0073010
1                                4000 - 4999 = ' 4000 - 4999 '
0074010
2                                5000 - 5999 = ' 5000 - 5999 '
```

```
0075010
3                 6000 - HIGH = ' 6000 AND OVER ' ;
NOTE : FORMAT INCGRP HAS BEEN OUTPUT .
3
0076010

NOTE   THE PROCEDURE FORMAT USED 0 02 CPU SECONDS AND 2054 K .

4                  DATA SAEED ;
0080000
5                  INPUT AGE 1 - 2 CITY$ 4 SEX$ 6 NAT$ 8 RPLACE $ 10 - 13 MARRD 15 OCCP 17
0090006
6                  NFAMLY 19 EDUC 21 INCOME 23 - 26 KDRG 28 RASON 30 ,
0091005
7                  FORMAT AGE AGROUP . ;
0100000
8                  FORMAT INCOME INCGRP . ;
0101010
9                 CARDS .
0110000

NOTE   THE DATA SET WORK   SAEED HAS 39 OBSERVATIONS AND 12 VARIABLES .
NOTE : THE DATA STATEMENT USED 0 06 CPU SECONDS AND 2763 K .

9
0110000
9
0154400
10                 PROC PRINT DATA = SAEED ;
,005 512

NOTE : THE PROCEDURE PRINT PRINTED PAGE 1
NOTE   THE PROCEDURE PRINT USED 0 04 CPU SECONDS AND 2857 K

11                 PROC SORT .
0104600
12                 BY CITY ;
0154700

NOTE : THE DATA SET WORK . SAEED HAS 39 OBSERVATIONS AND 12 VARIABLES .
NOTE . THE PROCEDURE SORT USED 0 02 CPU SECONDS AND 3042 K

15                 PROC FREQ .
0154800
15                 TABLES CITY .
0154900
16                 TABLES NAT * CITY ;
0155000
16                 TABLES AGE ;
0155100

NOTE : THE PROCEDURE FREQ PREQ PRINTED PAGE 2
NOTE   THE PROCEDURE FREQ USED 0 03 CPU SECONDS AND 3257 K

17                 PROC FREQ
0155208
18                 TABLES AGE * CITY .
```

```
00155308

NOTE : THE PROCEDURE FREQ PRINTED PAGE 3 .
NOTE . THE PROCEDURE FREQ USED 0 02 CPU SECONDS AND 2357 K .

69                PROC MEANS ;
00155408
70                VAR AGE .
00155508
71                BY CITY :
00155608

NOTE . THE PROCEDURE MEANS PRINTED PAGE 4 .
NOTE   THE PROCEDURE MEANS USED 0 02 SECONDS AND 3407 K .

72              PROC MEANS N MEAN STD VAR STDERR LCLM UCLM MAXDEC = 3 ,
00155709
73              VAR GAE ;
00155808
74              BY CITY .
00155908

NOTE . THE PROCEDURE MEANS PRINTED PAGE 5
NOTE   THE PROCEDURE MEANS USED 0 02 CPU SECONDS AND 3407 K

75                PROC MEANS ;
00156011
76                VAR INCOME .
00156111
77                BY CITY .
00156211

NOTE   THE PROCEDURE MEANS PRINTED PAGE 6
NOTE   THE PROCEDURE MEANS USED 0 01 CPU SECONDS AND 4307 K .

78                  PROC MEANS N MEAN STD VAR STDERR LCLM UCLM MAX-
DEC = 3 .
00156309
79                VAR INCOME .
00156409
80                BY CITY :
00156509

NOTE : THE PROCEDURE MEANS PRINTED PAGE 7
NOTE   THE PROCEDURE MEANS USED 0 02 CPU SECONDS AND 4307 K

81                PROC PRINT DATA = SAEED ,
00157011

NOTE . THE PROCEDURE PRINT PRINTED PAGE 8 .
NOTE   THE PROCEDURE PRINT USED 0 02 CPU SECONDS AND 3407 K
```

10 : 33 SUNDAY , JUNE 12 , 1994

CITY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	14	35.9	14	35.9
B	16	41.0	30	76.9
C	9	23.1	39	100.0

TABLE OF NAT BY CITY

NAT Frequency Percent Row Pct Col Pct	CITY A	B	C	Total
N	5 12.82 26.32 35.71	9 23.08 47.37 56.25	5 12.82 26.32 55.56	19 48.72
S	9 23.08 45.00 64.29	7 17.95 35.00 43.75	4 10.26 20.00 44.44	20 51.28
Total	14 35.90	16 41.03	9 23.08	39 100.00

AGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
15 - 19	10	25.6	10	25.6
20 - 24	15	38.5	25	64.1
25 - 29	5	12.8	30	76.9
30 - 34	9	23.1	39	100.0

TABLE OF AGE BY CITY

AGE	CITY			
Frequency Percent Row Pct Col Pct	A	B	C	Total
15 - 19	3 7.69 30.00 21.43	5 12.82 50.00 31.25	2 5.13 20.00 22 22	10 25.64
20 - 24	7 17.95 46.67 50.00	6 15.38 40.00 37.50	2 5.13 13.33 22.22	15 38.46
25 - 29	0 0.00 0.00 0.00	2 5.13 40.00 12.50	3 7.69 60.00 33.33	5 12.82
30 - 34	4 10.26 44.44 28.57	3 7.69 33.33 18.75	2 5.13 22.22 22.22	9 23.08
Total	14 35.90	16 41.03	9 23.08	39 100.00

ANALYSIS VARIABLE : AGE

---------------------- CITY = A ----------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
14	24.0000000	6.5044364	15.0000000	34.0000000

---------------------- CITY = B ----------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
16	23.0000000	5.9665736	15.0000000	33.0000000

---------------------- CITY = B ----------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
9	24.0000000	6.1327898	15.0000000	33.0000000

10 : 33 SUNDAY , JUNE 12 , 1994

ANALYSIS VARIABLE : AGE

CITY = A

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
14	24.000	6.504	42.308	1.738	20.244

UPPER 95.0% CLM
27.756

CITY = B

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
16	23.000	5.967	35.600	1.492	19.820

UPPER 95.0% CLM
26.179

CITY = C

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
9	24.111	6.133	37.611	2.044	19.397

UPPER 95.0% CLM
28.825

10 : 33 SUNDAY , JUNE 12 , 1994

ANALYSIS VARIABLE : INCOME

------------------- CITY = A -------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
14	4147.14	862.9326435	3300.00	6400.00

------------------- CITY = B -------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
16	4088.75	856.7837144	3400.00	6400.00

------------------- CITY = C -------------------

N	MEAN	STD DEV	MINIMUM	MAXIMUM
9	4122.22	603.6923425	3400.00	5300.00

ANALYSIS VARIABLE : INCOME

---------------------- CITY = A ----------------------

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
14	4147.143	862.933	744652.747	230.628	3648.900

UPPER 95.0% CLM
4645.385

---------------------- CITY = B ----------------------

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
16	4088.750	856.784	734078.333	214.196	3632.20

UPPER 95.0% CLM
4545.298

---------------------- CITY = C ----------------------

N	MEAN	STD DEV	VARIANCE	STD ERROR	LOWER 95.0% CLM
9	4122.222	603.692	364444.444	201.231	3658.103

UPPER 95.0% CLM
4586.261

10 : 34 SUNDAY , JUNE 12 , 1994

OBS	AGE	CITY	SEX	NAT	RPLACE	MARRD	OCCP	NFAMLY	EDUC	INCOME	KDRG	RASON
1	20	A	M	S	RIYD	1	1	5	4	3300	2	3
2	24	A	M	S	JEDH	2	2	4	2	4400	4	2
3	22	A	F	N	RIYD	3	3	5	4	3600	1	4
4	34	A	M	N	JEDH	3	2	3	3	4700	2	3
5	33	A	M	S	RIYD	4	1	2	3	4600	3	4
6	22	A	F	N	RIYD	4	3	5	2	4700	2	3
7	21	A	M	N	JEDH	3	4	5	4	3600	3	2
8	33	A	M	N	RIYD	4	2	6	2	4500	3	4
9	22	A	M	S	JEDH	1	1	4	4	3400	2	4
10	22	A	F	S	RIYD	2	4	5	3	4560	1	1
11	33	A	F	S	DAMM	3	3	4	3	3500	2	2
12	18	A	M	S	DAMM	2	4	3	2	3500	3	3
13	17	A	F	S	RIYD	3	3	5	2	6400	4	3
14	15	A	M	S	RIYD	2	2	6	1	3300	1	4
15	22	B	M	N	JEDH	1	1	4	4	3400	2	4
16	22	B	F	S	RIYD	2	4	5	3	4560	1	1
17	33	B	F	N	DAMM	3	3	4	3	3500	2	2
18	18	B	M	N	DAMM	2	4	3	2	3500	3	3
19	17	B	F	N	RIYD	3	3	5	2	6400	4	3
20	22	B	M	S	JEDH	1	1	4	4	3400	2	2
21	22	B	F	S	RIYD	2	4	5	3	4560	1	4
22	33	B	F	N	DAMM	3	3	4	3	3500	2	2
23	18	B	M	N	DAMM	2	4	3	2	3500	3	3
24	17	B	F	N	RIYD	3	3	5	2	4400	4	3
25	15	B	M	N	RIYD	2	2	6	1	5300	1	4
26	20	B	M	S	DAMM	4	2	4	1	4500	2	4
27	22	B	M	S	RIYD	3	4	6	2	3600	3	3
28	33	B	M	S	JEDH	3	3	4	2	4400	3	2
29	28	B	F	N	JEDH	3	2	5	3	3500	2	1
30	26	B	M	S	RIYD	2	1	5	4	3400	1	3
31	17	C	F	N	RIYD	3	3	5	2	4400	4	2
32	15	C	M	S	RIYD	2	2	6	1	5300	1	4
33	20	C	M	N	DAMM	4	2	4	1	4500	2	4
34	22	C	M	S	RIYD	3	4	6	2	3600	3	3
35	33	C	M	N	JEDH	3	3	4	2	4400	3	2
36	28	C	F	N	JEDH	3	2	5	3	3500	2	1
37	26	C	M	N	RIYD	2	1	5	4	3400	1	1
38	25	C	M	S	DAMM	2	2	4	3	4100	3	2
39	31	C	M	S	DAMM	3	3	3	4	3900	2	3

الملاحق

ملحق رقم (١) : جدول توزيع المنحنى الطبيعي .

ملحق رقم (٢) : جدول توزيع منحنى (ت) .

ملحق رقم (٣) : نموذج استمارة .

ملحق رقم (٤) : جداول الأرقام العشوائية .

ملحق رقم (٥) : إثبات بعض العلاقات والصيغ .



ملحق رقم (١)

Areas of a standard normal distribution
An entry in the table is the proportion under the entire curve that is between z = o and a positive value of z. Areas for negative values of z are obtained by symmetry .

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.5425	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4694
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	0.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997

PAUL G. HOEL : Basic Statistics For Business & Economics : المصدر : (ملحق رقم ١ ، وملحق رقم ٢)

ملحق رقم (٢)

Student's distribution
The first column lists the number of degrees of freedom (r). The headings of the other columns give probabilities (P) that t exceeds the entry value. Ues symmetry for negative t-values.

n	.10	.05	P .025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	1.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.516	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
00	1.282	1.645	1.960	2.326	2.576

ملحق رقم (٣)

المملكة العربية السعودية وزارة المالية والاقتصاد الوطني مصلحة الإحصاءات العامة التعداد الاقتصادي	**مشروع التعداد الاقتصادي** **استمارة تعداد المنشآت** **لعام ١٤١٥هـ (١٩٩٤م)**	هذه البيانات تعتبر سرية ولا تستخدم سوى للأغراض الإحصائية وفقاً للمرسوم الملكي ٢٣ الصادر في ١٣٧٩/١٢/٧هـ رقم الصفحة رقم تسلسل المنشأة

أولاً : البيانات الجغرافية والمميزة :

اسم الحي : ______ الرقم ١٠١ ______ رقم القطاع ١٠١ ______ رقم البلك ١٠٢ ______

اسم المنشأة ١٠٤ ______

عنوان المنشأة ١٠٥ اسم الشارع ______ رقم المبنى ______

الموقع بالتفصيل ______

رقم الهاتف ١٠٦ ______ صندوق البريد ١٠٧ ______ الرمز البريدي ١٠٨ ______

رقم السجل التجاري ١٠٩ ______ مكان الإصدار (المدينة) ______ رقمها ١١٠ ______

حالة العمل بالمنشأة (ضع علامة × في المربع المناسب) وأكتب الرقم في هذا المربع ١١١ ☐

١ ☐ عاملة ٢ ☐ مغلقة بصفة مؤقتة ٣ ☐ تحت التأسيس ٤ ☐ مغلقة بصفة دائمة

السنة التي بدأ فيها مزاولة النشاط ١١٢ ______

ثانياً : صفة المنشأة : (ضع علامة × في المربع المناسب)

١ ☐ منشأة فردية ليس لها فروع ٢ ☐ مركز رئيسي ٣ ☐ فرع أكتب الرقم في هذا المربع ٢٠١ ☐

إذا كان الموقع فرعاً يرجى تسجيل البيانات التالية

اسم المركز الرئيسي ______ المدينة ______ رقمها ٢٠٢ ______

اسم الحي ______ اسم الشارع ______

رقم الهاتف ٢٠٣ ______ صندوق البريد ٢٠٤ ______ الرمز البريدي ٢٠٥ ______

ثالثاً : الصفة القانونية : (ضع علامة × في المربع واكتب هذا الرقم في هذا المربع) ٣٠١ ☐

١ ☐ مؤسسة فردية ٢ ☐ شركة تضامن ٣ ☐ شركة توصية بالأسهم ٤ ☐ شركة ذات مسئولية محدودة ٥ ☐ شركة مساهمة ٦ ☐ حكومي ٧ ☐ أخرى

رابعاً : النشاط الاقتصادي الرئيسي : أكتب النشاط الاقتصادي الرئيسي للمنشأة بالتفصيل :

______ الرمز الرباعي للنشاط ٤٠١ ______

خامساً : الافراد المشتغلون خلال العام المالي السابق ١٩٩٤م أو (١٤١٣ - ١٤١٤هـ)

ملاحظة :
الأفراد المشتغلون تشمل جميع العاملين بالمنشأة من الملاك والإداريين والمتخصصين وعمال الإنتاج وغيرهم

سعوديون ٥٠١	غير سعوديين ٥٠٢	الجملة ٥٠٣

سادساً : جملة النفقات والايرادات خلال العام المالي السابق ١٩٩٣م (١٤١٣ - ١٤١٤هـ)

		الرمز	الاف الريالات
الرواتب والأجور للمشتغلين لعام ١٩٩٣م (١٤١٣ - ١٤١٤هـ)	للسعوديين	٦٠١	
	لغير السعوديين	٦٠٢	
	جملة الرواتب والأجور	٦٠٣	
النفقات (بأستثناء الرواتب والأجور)		٦٠٤	
جملة النفقات (البند ٦٠٣ ، البند ٦٠٤)		٦٠٥	
جملة الإيرادات		٦٠٦	

اسم معطي البيانات ______ رقم الهاتف ______

ملاحظات
....
....

توقيع المفتش بالمراجعة

التاريخ

ملحق رقم (٤)

Random numbers

31 75 15 72 60	68 98 00 53 39	15 47 04 83 55	88 65 12 25 96	03 15 21 91 21
88 49 29 93 82	14 45 40 45 04	20 09 49 89 77	74 84 39 34 13	22 10 97 85 08
30 93 44 77 44	07 48 18 38 28	73 78 80 65 33	28 59 72 04 05	94 20 52 03 80
22 88 84 88 93	27 49 99 87 48	60 53 04 51 26	74 02 28 46 17	82 03 71 02 68
78 21 21 69 93	35 90 29 13 86	44 37 21 54 86	65 74 11 40 14	87 48 13 72 20
41 84 98 45 47	46 85 05 23 26	34 67 75 83 00	74 91 06 43 45	19 32 58 15 49
46 35 23 30 49	69 24 89 34 60	45 30 50 75 21	61 31 83 18 55	14 41 37 09 51
11 08 79 62 94	14 01 33 17 92	59 74 76 72 77	76 50 33 45 13	39 66 37 75 44
52 70 10 83 37	56 30 38 73 15	16 52 06 96 76	11 65 49 98 93	02 18 16 81 61
57 27 53 68 98	81 30 44 85 85	68 65 22 73 76	92 85 25 58 66	88 44 80 35 84
20 85 77 31 56	70 28 42 43 26	79 37 59 52 20	01 15 96 32 67	10 62 24 83 91
15 63 38 49 24	90 41 59 36 14	33 52 12 66 65	55 82 34 76 41	86 22 53 17 04
92 69 44 82 97	39 90 40 21 15	59 58 94 90 67	66 82 14 15 75	49 76 70 40 37
77 61 31 90 19	88 15 20 00 80	20 55 49 14 09	96 27 74 82 57	50 81 60 76 16
38 68 83 24 86	45 13 46 35 45	59 40 47 20 59	43 94 75 16 80	43 85 25 96 93
25 16 30 18 89	70 01 41 50 21	41 29 06 73 12	71 85 71 59 57	68 97 11 14 03
65 25 10 76 29	37 23 93 32 95	05 87 00 11 19	92 78 42 63 40	18 47 76 56 22
36 81 54 36 25	18 63 73 75 09	82 44 49 90 05	04 92 17 37 01	14 70 79 39 97
64 39 71 16 92	05 32 78 21 62	20 24 78 17 59	45 19 72 53 32	83 74 52 25 67
04 51 52 56 24	95 09 66 79 46	48 46 08 55 58	15 19 11 87 82	16 93 03 33 61
15 88 09 22 61	17 29 28 81 90	61 78 14 88 98	92 52 52 12 83	88 58 16 00 98
71 92 60 08 19	59 14 40 02 24	30 57 09 01 94	18 32 90 69 99	26 85 71 92 39
64 42 52 81 08	16 55 41 60 16	00 04 28 32 29	10 33 33 61 68	65 61 79 48 34
79 78 22 39 24	49 44 03 04 32	81 07 73 15 43	95 21 66 48 65	13 65 85 10 81
35 33 77 45 38	44 55 36 46 72	90 96 04 18 49	93 86 54 46 08	93 17 63 48 51
05 24 92 93 29	19 71 59 40 82	14 73 88 66 67	43 70 86 63 54	93 69 22 55 27
56 46 39 93 80	38 79 38 57 74	19 05 61 39 39	46 06 22 76 47	66 14 66 32 10
96 29 63 31 21	54 19 63 41 08	75 81 48 59 86	71 17 11 51 02	28 99 26 31 65
98 38 03 62 69	60 01 40 72 01	62 44 84 63 85	42 17 58 83 50	46 18 24 91 26
52 56 76 43 50	16 31 55 39 69	80 39 58 11 14	54 35 86 45 78	47 26 91 57 47
78 49 89 08 30	25 95 59 92 36	43 28 69 10 64	99 96 99 51 44	64 42 47 73 77
49 55 32 42 41	08 15 08 95 35	08 70 39 10 41	77 32 38 10 79	45 12 79 63 86
32 15 10 70 75	83 15 51 02 52	73 10 08 86 18	23 89 18 74 18	45 41 72 02 68
11 31 45 03 63	26 86 02 77 99	49 41 68 35 34	19 18 70 80 59	76 67 70 21 10
12 36 47 12 10	87 05 25 02 41	90 78 59 78 89	81 39 95 81 30	64 43 90 58 14
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 62 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 05 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08

Random numbers (continued)

77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
60 31 14 28 24	37 30 14 26 78	45 99 04 32 42	17 37 45 20 03	70 70 77 02 14
49 73 97 14 84	92 00 39 80 86	76 66 87 32 09	59 20 21 19 73	02 90 23 32 50
78 62 65 15 94	16 45 39 46 14	39 01 49 70 66	83 01 20 98 32	25 57 17 76 28
66 69 21 39 86	99 83 70 05 82	81 23 24 49 87	09 50 49 64 12	90 19 37 95 68
44 07 12 80 91	07 36 29 77 03	76 44 74 25 37	98 52 49 78 31	65 70 40 95 14
41 46 88 51 49	49 55 41 79 94	14 92 43 96 50	95 29 40 05 56	70 48 10 69 05
94 55 93 75 59	49 67 85 31 19	70 31 20 56 82	66 98 63 40 99	74 47 42 07 40
41 61 57 03 60	64 11 45 86 60	90 85 06 46 18	80 62 05 17 90	11 43 63 80 72
50 27 39 31 13	41 79 48 68 61	24 78 18 96 83	55 41 18 56 67	77 53 59 98 92
41 39 68 05 04	90 67 00 82 89	40 90 20 50 69	95 08 30 67 83	28 10 25 78 16
25 80 72 42 60	71 52 97 89 20	72 68 20 73 85	90 72 65 71 66	98 88 40 85 83
06 17 09 79 65	88 30 29 80 41	21 44 34 18 08	68 98 48 36 20	89 74 79 88 82
60 80 85 44 44	74 41 28 11 05	01 17 62 88 38	36 42 11 64 89	18 05 95 10 61
80 94 04 48 93	10 40 83 62 22	80 58 27 19 44	92 63 84 03 33	67 05 41 60 67
19 51 69 01 20	46 75 97 16 43	13 17 75 52 92	21 03 68 28 08	77 50 19 74 27
49 38 65 44 80	23 60 42 35 54	21 78 54 11 01	91 17 81 01 74	29 42 09 04 38
06 31 28 89 40	15 99 26 93 21	47 45 86 48 09	98 18 98 18 51	29 65 18 42 15
60 94 20 03 07	11 89 79 56 74	40 40 56 80 32	96 71 75 42 44	10 70 14 13 93
92 32 99 89 32	78 28 44 63 47	71 20 99 20 61	39 44 89 31 36	25 72 20 85 64
77 93 66 35 74	31 38 45 19 24	85 56 12 96 71	58 13 71 78 20	22 75 13 65 18
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93	28 97 66 62 52
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90	09 81 59 31 46
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92	54 13 05 51 60
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01	14 97 44 03 44
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51	43 66 77 08 83
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77	90 71 22 67 69
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23	08 81 64 74 49
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98	16 43 59 15 29
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22	26 65 59 08 02
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37	41 32 64 43 44
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31	96 24 04 36 42
24 63 73 87 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01	03 74 28 38 73
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85	51 97 23 78 67
16 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52	54 84 65 47 59
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79	65 13 00 48 60
94 01 54 68 74	32 44 44 82 77	59 82 09 61 63	64 65 42 58 43	41 14 54 28 20
74 10 88 82 22	88 57 07 40 15	25 70 49 10 35	01 75 51 47 50	48 96 83 86 03
62 88 08 78 73	95 16 05 92 21	22 30 49 03 14	72 87 71 73 34	39 28 30 41 49
11 74 81 21 02	80 58 04 18 67	17 71 05 96 21	06 55 40 78 50	73 95 07 95 52
17 94 40 56 00	60 47 80 33 43	25 85 25 89 05	57 21 63 96 18	49 85 69 93 26

Random numbers (continued)

66 06 74 27 92	95 04 35 26 80	46 78 05 64 87	09 97 15 94 81	37 00 62 21 86
54 24 49 10 30	45 54 77 08 18	59 84 99 61 69	61 45 92 16 47	87 41 71 71 98
30 94 55 75 89	31 73 25 72 60	47 67 00 76 54	46 37 62 53 66	94 74 64 95 80
69 17 03 74 03	86 99 59 03 07	94 30 47 18 03	26 82 50 55 11	12 45 99 13 14
08 34 58 89 75	35 84 18 57 71	08 10 55 99 87	87 11 22 14 76	14 71 37 11 81
27 76 74 35 84	85 30 18 89 77	29 49 06 97 14	73 03 54 12 07	74 69 90 93 10
13 02 51 43 38	54 06 61 52 43	47 72 46 67 33	47 43 14 39 05	31 04 85 66 99
80 21 73 62 92	98 52 52 43 35	24 43 22 48 96	43 27 75 88 74	11 46 61 60 82
10 87 56 20 04	90 39 16 11 05	57 41 10 63 68	53 85 63 07 43	08 67 08 47 41
54 12 75 73 26	26 62 91 90 87	24 47 28 87 79	30 54 02 78 86	61 73 27 54 54
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22	93 17 49 39 72
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00	71 14 84 36 43
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61	62 32 71 84 23
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60	81 60 41 88 80
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69	85 64 44 72 77
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49	65 58 44 96 98
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05	40 03 03 74 38
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47	15 50 12 95 78
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41	03 85 65 45 52
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56	64 69 11 92 02
07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38	04 71 36 69 94
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39	61 21 20 64 55
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83	92 30 15 04 98
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39	06 41 01 93 62
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51	31 02 47 31 67
59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56	04 11 10 84 08
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07	95 95 44 99 53
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00	05 46 26 92 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76	96 29 99 08 36
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11	97 34 13 03 58
38 10 17 77 56	11 65 71 38 97	95 88 95 70 67	47 64 81 38 85	70 66 99 34 06
39 64 16 94 57	91 33 92 25 02	92 61 38 97 19	11 94 75 62 03	19 32 42 05 04
84 05 44 04 55	99 39 66 36 80	67 66 76 06 31	69 18 19 68 45	38 52 51 16 00
47 46 80 35 77	57 64 96 32 66	24 70 07 15 94	14 00 42 31 53	69 24 90 57 47
43 32 13 13 70	28 97 72 38 96	76 47 96 85 62	62 34 20 75 89	08 89 90 59 85
64 28 16 18 26	18 55 56 49 37	13 17 33 33 65	78 85 11 64 99	87 06 41 30 75
66 48 77 04 95	32 35 00 29 85	86 71 63 87 46	26 31 37 74 63	55 38 77 26 81
72 46 13 32 30	21 52 95 34 24	92 58 10 22 62	78 43 86 62 76	18 39 67 35 38
21 03 29 10 50	13 05 81 62 18	12 47 05 65 00	15 29 27 61 39	59 52 65 21 13
95 36 26 70 11	06 65 11 61 36	01 01 60 08 57	55 01 85 63 74	35 82 47 17 08
40 71 29 73 80	10 40 45 54 52	34 03 06 07 26	75 21 11 02 71	36 63 36 84 24
58 27 56 17 64	97 58 65 47 16	50 25 94 63 45	87 19 54 60 92	26 78 76 09 39
89 51 41 17 88	68 22 42 34 17	73 95 97 61 45	30 34 24 02 77	11 04 97 20 49
15 47 25 06 69	48 13 93 67 32	46 87 43 70 88	73 46 50 98 19	58 86 93 52 20
12 12 08 61 24	51 24 74 43 02	60 88 35 21 09	21 43 73 67 88	49 22 67 78 37

Random numbers (continued)

03 99 11 04 61	93 71 61 68 94	66 08 32 46 53	84 60 95 82 32	88 61 81 91 61
38 55 59 55 54	32 88 65 97 80	08 35 65 08 60	29 73 54 77 62	71 29 92 38 53
17 54 67 37 04	92 05 24 62 15	55 12 12 92 81	59 07 60 79 36	27 95 45 89 09
32 64 35 28 61	95 81 90 68 31	00 91 19 89 36	76 35 59 37 79	80 86 30 05 14
69 57 26 87 77	39 51 03 59 05	14 06 04 06 19	29 54 96 96 16	33 56 46 07 80
24 12 26 65 91	27 69 90 64 94	14 84 54 66 72	61 95 87 71 00	90 89 97 57 54
61 19 63 02 31	92 96 26 17 73	41 83 95 53 82	17 26 77 09 43	78 03 87 02 67
30 53 22 17 04	10 27 41 22 02	39 68 52 33 09	10 06 16 86 29	55 98 66 64 85
03 78 89 75 99	75 86 72 07 17	74 41 65 31 66	35 20 83 33 74	87 53 90 88 23
48 22 86 33 79	85 78 43 76 19	53 15 26 74 33	35 66 35 29 72	16 81 86 03 11
60 36 59 46 53	35 07 53 39 49	42 61 42 92 97	01 91 82 83 16	98 95 37 32 31
83 79 94 24 02	56 62 33 44 42	34 99 44 13 74	70 07 11 47 36	09 95 81 80 65
32 96 00 74 05	36 40 98 32 32	99 38 54 16 00	11 13 30 75 86	15 91 70 62 53
19 32 25 38 45	57 62 05 26 06	66 49 76 86 46	78 13 86 65 59	19 64 09 94 13
11 22 09 47 47	07 39 93 74 08	48 50 92 39 29	27 48 24 54 76	85 24 43 51 59
21 44 58 27 93	24 83 19 32 41	14 19 97 62 68	70 88 36 80 02	03 82 91 74 43
72 51 37 64 00	52 22 59 23 48	62 30 89 84 81	29 74 43 31 65	33 14 16 10 20
71 47 94 50 27	76 16 05 74 11	13 78 01 36 32	52 30 87 77 62	88 87 43 36 97
83 21 05 14 66	09 08 85 03 95	26 74 30 53 06	21 70 67 00 01	99 43 98 07 67
68 74 99 51 48	94 89 77 86 36	96 75 00 90 24	94 53 89 11 43	96 69 36 18 86
05 18 47 57 63	47 07 58 81 58	05 31 35 34 39	14 90 80 88 30	60 09 62 15 51
13 65 16 25 46	96 89 22 52 40	47 51 15 84 83	87 34 27 88 18	07 85 53 92 69
00 56 62 12 20	00 29 22 40 69	25 07 22 95 19	52 54 85 40 91	21 28 22 12 96
05 95 81 76 95	58 07 26 89 90	60 32 99 59 55	71 58 66 34 17	35 94 76 78 07
57 62 16 45 47	46 85 03 79 81	38 52 70 90 37	64 75 60 33 24	04 98 68 36 66
09 28 22 58 44	79 13 97 84 35	35 42 84 35 61	69 79 96 33 14	12 99 19 35 16
23 39 49 42 06	93 43 23 78 36	94 91 92 68 46	02 55 57 44 10	94 91 54 81 99
05 28 03 74 70	93 62 20 43 45	15 09 21 95 10	18 09 41 66 13	78 23 45 00 01
95 49 19 79 76	38 30 63 21 92	82 63 95 46 24	72 43 49 26 06	23 19 17 46 93
78 52 10 01 04	18 24 87 55 83	90 32 65 07 85	54 03 46 62 51	35 77 41 46 92
96 34 54 45 79	85 93 24 40 53	75 70 42 08 40	86 58 38 39 44	52 45 67 37 66
77 96 33 11 51	32 36 49 16 91	47 35 74 03 38	23 43 52 40 65	08 45 89 53 66
07 52 01 12 94	23 23 80 17 48	41 69 06 73 28	54 81 43 77 77	10 05 74 23 32
38 42 30 23 09	70 70 38 57 36	46 14 81 42 48	29 23 61 21 52	05 08 86 58 52
02 46 36 55 33	21 19 96 05 55	33 92 80 18 17	07 39 68 92 15	30 72 22 21 02
38 76 16 08 73	43 25 38 41 45	60 83 32 59 83	01 29 14 13 49	20 36 80 71 26
14 38 70 63 45	80 85 40 92 79	43 52 90 63 18	38 38 47 47 61	41 19 63 74 80
51 32 19 22 46	80 08 87 70 74	88 72 25 67 36	66 16 44 94 31	66 91 93 16 78
72 47 20 00 08	80 89 01 80 02	94 81 33 19 00	54 15 58 34 36	35 35 25 41 31
05 46 65 53 06	93 12 81 84 64	74 45 79 05 61	72 84 81 18 34	79 98 26 84 16
39 52 87 24 84	82 47 42 55 93	48 54 53 52 47	18 61 91 36 74	18 61 11 92 41
81 61 61 87 11	53 34 24 42 76	75 12 21 17 24	74 62 77 37 07	58 31 91 59 97
07 58 61 61 20	82 64 12 28 20	92 90 41 31 41	23 39 21 97 63	61 19 96 79 40
90 76 70 42 35	13 57 41 72 00	69 90 26 37 42	78 46 42 25 01	18 62 79 08 72
40 18 82 81 93	29 59 28 86 27	94 97 21 15 98	62 09 53 67 87	00 44 15 89 97

Random numbers (continued)

34 41 48 21 57	86 88 75 50 87	19 15 20 00 23	12 30 28 07 83	32 62 46 86 91
63 43 97 53 63	44 98 91 68 22	36 02 40 08 67	76 37 84 16 05	65 96 17 34 88
67 04 90 90 70	93 39 94 55 47	94 45 87 42 84	05 04 14 98 07	20 28 83 40 60
79 49 50 41 46	52 16 29 02 86	54 15 83 42 43	46 97 83 54 82	59 36 29 59 38
91 70 43 05 52	04 73 72 10 31	75 05 19 30 29	47 66 56 43 82	99 78 29 34 78
19 61 27 84 30	11 66 19 47 70	77 60 36 56 69	86 86 81 26 65	30 01 27 59 89
39 14 17 74 00	28 00 06 42 38	73 25 87 17 94	31 34 02 62 56	66 45 33 70 16
64 75 68 04 57	08 74 71 28 36	03 46 95 06 78	03 27 44 34 23	66 67 78 25 56
92 90 15 18 78	56 44 12 29 98	29 71 83 84 47	06 54 32 53 11	07 56 55 37 71
03 55 19 00 70	09 48 39 40 50	45 93 81 81 35	36 90 84 33 21	11 07 35 18 03
98 88 46 62 09	06 83 05 36 56	14 66 35 63 46	71 43 00 49 09	19 81 80 57 07
27 36 98 68 82	53 47 30 75 41	53 63 37 08 63	03 74 81 28 22	19 36 04 90 88
59 06 67 59 74	63 33 52 04 83	43 51 43 74 81	58 27 82 69 67	49 32 54 39 51
91 64 79 37 83	64 16 94 90 22	98 58 80 94 95	49 82 95 90 68	38 83 10 48 38
83 60 59 24 19	39 54 20 77 72	71 56 87 56 73	35 18 58 97 59	44 90 17 42 91
24 89 58 85 30	70 77 43 54 39	46 75 87 04 72	70 20 79 26 75	91 62 36 12 75
35 72 02 65 56	95 59 62 00 94	73 75 08 57 88	43 26 40 17 03	46 83 36 52 48
14 14 15 34 10	38 64 90 63 43	57 25 66 13 42	72 70 97 53 18	90 37 93 75 62
27 41 67 56 70	92 17 67 25 35	93 11 95 60 77	06 88 61 82 44	92 34 43 13 74
82 07 10 74 29	81 74 74 77 49	40 74 45 69 74	23 33 68 88 21	53 84 11 05 36

أخذت بيانات هذه الجداول من كتاب :

PAUL G. HOEL : Basic Statistics For Business & Economics.

ملحق رقم (٥)
ملحق رقم (٥ ـ ١)

المعاينة العشوائية البسيطة :

– إثبات أن تباين تقدير متوسط المجتمع فى حال السحب مع الإعادة يساوى :

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

$$V(x_i) = E\{x_i - \bar{X}\}^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 P(x_i = X_i)$$

لدينا

$$= \sum_{i=1}^{N} \frac{1}{N} \cdot (x_i - \bar{X})^2 = \sigma^2$$

إذا رمزنا لمجموع مفردات العينة بالرمز x أى $x = \sum_{i=1}^{n} x_i$ يكون لدينا :

$$V(x) = V\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} V(x_i) = n\,V(x_i) = n\sigma^2$$

وبالتالى :

$$V(\bar{x}) = V\left(\frac{x}{n}\right) = \frac{1}{n^2}\,V(x) = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n}$$

وهو المطلوب . ويكون فى هذه الحالة الخطأ المعيارى .

$$\sigma_{\bar{x}} = \sqrt{V(\bar{x})} = \frac{\sigma}{\sqrt{n}}$$

– إثبات أن تباين تقدير متوسط المجتمع فى حال السحب مع عدم الإعادة يساوى :

$$V(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{S^2}{n}(1-f)$$

حيث $f = \frac{n}{N}$

إذا كان $x=\sum_{i=1}^{n} x_i$ فإن تباين مجموع مشاهدات العينة :

$$V\{x\} = E\left\{\left[\sum_{i=1}^{n} x_i - n\bar{X}\right]^2\right\}$$

حيث $E(x) = n\bar{X}$

$$= E\left\{\sum_{k=1}^{n}\sum_{i=1}^{n}\{x_i-\bar{x}\}\{x_k-\bar{x}\}\right\} + E\left\{\sum_{i=1}^{n}\{x_i-\bar{x}\}\right\}^2$$

$$= E\left\{\sum_{i}^{n}\{x_i-\bar{x}\}^2\right\} + \left\{\left\{\sum_{i\neq k}^{n}\{x_i-\bar{X}\}\{x_k-\bar{X}\}\right\} P\{x = x_{i}, x_j\}\right.$$

ونظرا لعدم وجود استقلال تام ، فإن التغاير بين جميع أزواج المفـــردات (k,i) المختارة لايتلاشى ، وذلك في حالة السحب مع عدم الإعادة ، لذا فإن القيمة المتوقعة لكل (n (n - 1 من أزواج تغاير العينة هى القيمة المتوسطة من بين (N (N - 1 من أزواج تغاير المجتمع ٠ وبشكل مشابه فإن القيمة المتوقعة لكل من (n) من قيم التباين هى تباين مفردات المجتمع ٠ لذا نجد أن :

$$VAR\{x\} = \frac{n}{N}\sum_{i=1}^{N}\{x_i-\bar{X}\}^2 + \frac{n\{n-1\}}{N\{N-1\}}\left\{\sum_{i\neq k}^{N}\{x_i-\bar{X}\}\{x_k-\bar{X}\}\right\}$$

$$= n\sigma^2 + \frac{n}{N}\frac{(n-1)}{(N-1)}\left\{\left[\sum_{i=1}^{N}(x_i-\bar{x})\right]^2 - \sum_{i=1}^{N}(x_i-\bar{X})^2\right\}$$

$$= n\sigma^2 + \frac{n}{N}\frac{(n-1)}{(N-1)}(0 - N\sigma^2)$$

$$= n\sigma^2 + \frac{n}{N}\frac{(n-1)}{(N-1)}N\sigma^2 = \frac{N-n}{N-1}n\sigma^2$$

وحيث إن $\sigma^2 = \frac{(N-1)}{N}S^2$ نجد :

$$VAR\{x\} = nS^2\,\frac{N-n}{N} = nS^2\,(1-f)$$

حيث $f = \frac{n}{N}$

إن

$$V\{\bar{x}\} = Var\left(\frac{x}{n}\right) = \frac{1}{n^2} \; Var(x) = \frac{1}{n^2} \, n \, S^2 \;\; (1 - f)$$

$$= \frac{1}{n} S^2 \;\; (1 - f)$$

وبالتالي يكون الخطأ المعيارى للعينة العشوائية البسيطة فى حال السحب مع عدم الإعادة

$$\sigma_{\bar{x}} = \sqrt{VAR\{\bar{x}\}} = \frac{S}{\sqrt{n}} \quad \sqrt{1 - f}$$

ملحق رقم (٥ ـ ٣)

إثبات أن مقدر تباين المجتمع $\hat{V}(p_{st})$ هو مقدر غير متحيز لتباين تقدير نسبة المجتمع $V(p_{st})$ فى العينة الطبقية العشوائية ، أى نريد إثبات أن :

$$E\left\{\hat{V}\left(p_{st}\right)\right\} = V\left(p'_{st}\right)$$

$$= \frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h^2}{n_h}P_h\ Q_h$$

(وذلك عندما يكون معامل تصحيح المجتمع المحدود مساوياً للواحد)

نعلم أن :

$$E\left\{\hat{V}\left(p_{st}\right)\right\} = E\left\{\frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h^2}{n_h}p_h\ q_h\right\}$$

$$=\frac{1}{N^2}E\left(\frac{N_1^2}{n_1}p_1\ q_1+\frac{N_h^2}{n_2}p_2\ q_2+\ldots\ldots+\frac{N_L^2}{n_L}p_L\ q_L\right)$$

$$=\frac{1}{N^2}\left[\frac{N_1^2}{n_1}E(P_1)E(q_1)+\frac{N_2^2}{n_2}E(p_2)E(q_2)+\ldots\ldots+\frac{N_L^2}{n_L}E(P_L)E(q_L)\right]\ldots\ldots(2)$$

وأن توقع تقدير نسبة المجتمع للطبقة (h) يساوى :

$$E\left(p_h\right) = E\left\{\sum_{i=1}^{n_h}x_{hi}\right\}$$

$$=\frac{1}{n_h}\sum_{i=1}^{n_h}E\left(x_{hi}\right)=\frac{1}{n_h}n_h P_h = P_h$$

وبالطريقة نفسها نجد أن $E\ (q_h) = Q_h$

لأن $q_h = 1\text{-}P_h$ وتساوى

$$q_h = 1 - \sum_{i=1}^{n_h}x_{hi}$$

أى أن $E\ (q_h)\ 1-P_h = Q_h$

وبالتبديل فى الصيغة (2) نجد أن :

$$E\{\hat{V}\ (P_{st})\} = \frac{1}{N^2}\left[\frac{N_1^2}{n_1}\ P_1\ Q_1 + \frac{N_2^2}{n_2}\ P_2\ \ Q_2 + \ldots\ldots + \frac{N_L^2}{n_L} P_L\ \ Q_L\right]$$

$$- \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h^2}{n_h} P_h Q_h$$

وهو المطلوب .

ملحق رقم (٥-ب ٣)

- استخراج الصيغة المستخدمة لتحديد حجم العينة الطبقية لدينا حد خطأ التقدير الذي نقبله .

$$\beta = Z\sqrt{\hat{V}\{\bar{x}_{st}\}}$$

ومنه نجد أن $\frac{\beta^2}{Z^2} = \hat{V}(\bar{x}_{st}) = D$

حيث رمزناً لهذا الكسر بالرمز D أى أن :

$$D = \hat{V}(\bar{x}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h - n_h}{N_h}\;\frac{N_h^2 \quad s_h^2}{n_h}$$

ومنه

$$N^2 D = \sum_{h=1}^{L}\frac{N_h^2}{n_h}\,s_h^2 - \sum_{h=1}^{L} N_h\, s_h^2$$

وإذا رمزنا بالنسبة لحجم الطبقة (h) إلى إجمالى حجم العينة بالرمز w_h يكون $w_h = \frac{n_h}{n}$

ويساوى أيضاً $w_h = \frac{N_h}{N}$

ومنه نجد أن $n_h = n w_h$

وبالتالى يمكننا القول إن

$$N^2 D + \sum_{h=1}^{L} N_h\, s_h^2 = \sum_{h=1}^{L}\frac{N_h^2\, s_h^2}{n\;\; w_h}$$

$$= \frac{1}{n}\sum_{h=1}^{L}\frac{N_h^2 \quad s_h^2}{w_h}$$

ومنه نجد أن حجم العينة اللازم لتقدير متوسط المجتمع يساوي :

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2 \ s_h^2}{w_h}}{N^2 D + \sum_{h=1}^{L} N_h \ s_h^2}$$

حيث $D = \frac{\beta^2}{Z^2}$ و $w_h = \frac{n_h}{n} = \frac{N_h}{N}$

وهو المطلوب .

– استخراج حجم العينة للطبقة (h) للتوزيع الأمثل :

نعلم أن تباين تقدير وسطي المجتمع للعينة الطبقية يساوى :

$$V(\overline{x}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2}{n_h} S_h^2 \qquad \ldots\ldots \{1\}$$

ولدينا دالة تكاليف المعاينة بافتراض أن التكاليف مقسمة على جميع الطبقات أي

$$C = \sum_{h=1}^{L} n_h \ C_h \qquad \ldots\ldots \{2\}$$

حيث C_h هي تكلفة الوحدة في الطبقة (h) .

المطلوب تحديد حجم (n_h) بحيث تكون التكاليف أقل ما يمكن .

باستخدام دالة لاغرانج Lagrange Function نجد أن :

$$\varphi = \frac{1}{N^2} \sum_{h=1}^{L} \frac{N_h - n_h}{N_h} \frac{N_h^2 \ S_h^2}{n_h} + \lambda\Big(\sum_{h=1} n_h C_h - C\Big)$$

ولجعل هذه الدالة أقل ما يمكن نأخذ تفاضلها بالنسبة لـ n_h ونساوي الناتج بالصفر أي أن

$$\frac{\partial \ \phi}{\partial \ n_h} = -\frac{N_h^2 \ S_h^2}{N^2 \ n_h^2} + \lambda \ C_h = 0$$

ومنه

$$n_h = \frac{1}{\sqrt{\lambda}} \frac{N_h S_h}{N \sqrt{C_h}} \qquad \{3\}$$

وبأخذ المجموع لجميع الطبقات نجد أن .

$$\sum_{h=1}^{L} n_h = n = \frac{1}{\sqrt{\lambda}} \sum_{h=1}^{L} \frac{N_h S_h}{N \sqrt{C_h}}$$

ومنه نجد أن :

$$\sqrt{\lambda} = \frac{1}{n} \frac{1}{N} \sum_{h=1}^{L} \frac{N_h S_h}{\sqrt{C_h}}$$

وبالتعويض فى {3} نجد أن :

$$n_h = \frac{N_h C_h / \sqrt{C_h}}{\sum_{h-1}^{L} \{N_n S_h\} / \sqrt{C_h}} \, n$$

وهو المطلوب إثباته .

ملحق رقم (٥ ـ ٤)

- إثبات أن

$$V\{\overline{x}_{sy}\} = \frac{N-1}{N} S^2 - K\frac{\{n-1\}}{N} S_w^2$$

لدينا

$$V\{\overline{x}_{sy}\} = \frac{1}{K}\sum_{i=1}^{k}\{\overline{x}_i - \mu\}^2$$

إن مجموع مربعات انحرافات أوساط العينات عن متوسط المجتمع يساوى

$$\sum_{i=1}^{k}\{\overline{x}_i - \mu\}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\{x_{ij} - \mu\}^2$$

$$= \sum_{i}^{k}\sum_{j}^{n}\left|\{x_{ij} - \overline{x}_i\} + \{\overline{x}_i - \mu\}\right|^2$$

$$= \sum_{i}^{k}\sum_{j}^{n}\{x_{ij} - \overline{x}_i\}^2 + \sum_{i}^{k}\sum_{j}^{n}\{\overline{x}_i - \mu\}^2$$

$$= \sum_{i}^{k}\sum_{j}^{n}(x_{ij} - \overline{x}_i)^2 + n\sum_{i}^{k}(\overline{x}_i - \mu)^2$$

$$= K\{n-1\} S_w^2 + nk\, V\{\overline{x}_{sy}\}$$

حيث $\sum_{i}^{K}\sum_{j}^{n}(x_{ij} - \overline{x}_i)$ هو مجموع مربعات انحرافات القيم عن الوسط الحسابى لجميع العينات الممكنة .

ونعلم أن :

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\{x_{ij} - \mu\}^2 = \{N-1\} S^2$$

لذا نجد أن :

$$\{N - 1\} S^2 = K \{n - 1\} S_w^2 + n K V\{\overline{x}_{sy}\}$$

إن nK = N لذا نجد أن :

$$V\{\overline{x}_{sy}\} = \frac{N-1}{N} S^2 - K \frac{\{n-1\}}{N} S_w^2$$

وهوالمطلوب إثباته .

- إثبات أن

$$V\{\overline{x}_{sy}\} = \frac{N-1}{N} S^2 - \frac{1}{N} \sum_{j=1}^{n} \{x_{ij} - \overline{x}_i\}^2$$

حيث

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij} - \mu\}^2$$

نعلم أن :

$$V\{\overline{x}_{Sy}\} = \frac{1}{K} \sum_{i=1}^{k} \{\overline{x}_i - \mu\}^2$$

وأن

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij} - \mu\}^2 = \sum_{i}^{k} \sum_{j}^{n} \left[\{x_{ij} - \overline{x}_i\} + \{\overline{x}_i - \mu\}\right]^2$$

$$= \sum_{i}^{k} \sum_{j}^{n} \{x_{ij} - \overline{x}_i\}^2 + \sum_{i}^{k} \sum_{j}^{n} \{\overline{x}_i - \mu\}^2$$

$$+2 \sum_{i}^{k} \sum_{j}^{n} \{x_{ij} - \overline{x}_i\} . \{\overline{x}_i - \mu\}$$

$$= \sum_{i}^{k} \sum_{j}^{n} \{x_{ij} - \overline{x}_i\}^2 + n \sum_{i}^{k} \{\overline{x}_i - \mu\}^2$$

$$+ 2 \sum_{i}^{k} \{\overline{x}_i - \mu\} \sum_{j}^{n} \{x_{ij} - \overline{x}_i\}$$

إن المقدار الأخير يساوى الصفر .

ويمكننا القول إن :

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij}-\mu\}^2 = \sum_{i}^{k} \sum_{j}^{n} \{x_{ij}-\bar{x}_i\}^2 + n \sum_{i=1}^{k} \{\bar{x}_i-\mu\}^2$$

ومما سبق نجد أن :

$$K\, V\{\bar{x}_{sy}\} = \sum_{i=1}^{k} \{\bar{x}_i-\mu\}^2$$

وبالتبديل فى الحد الثانى من الطرف الأيمن للصيغة الأخيرة نجد أن

$$nk\, V\{\bar{x}_{sy}\} = \sum_{i}^{k} \sum_{j}^{n} \{x_{ij}-\mu\}^2 \quad \sum_{i}^{k} \sum_{j}^{n} \{x_{ij}-\bar{x}_i\}^2$$

أى أن :

$$V\{\bar{x}_{sy}\} = \frac{1}{nk} \sum_{i}^{k} \sum_{j}^{n} \{x_{ij} \;\; \mu\}^2 \quad \frac{1}{nk} \sum_{i}^{k} \sum_{j}^{n} \{x_{ij} - \bar{x}_i\}^2$$

ونعلم أن :

$$\sum_{i}^{k} \sum_{j}^{n} \{x_{ij}-\mu\}^2 = \{N-1\}S^2$$

وأن : $N = nk$

لذا نجد أن تباين متوسط المجتمع للعينة المنتظمة هو :

$$V\{\bar{x}_{sy}\} = \frac{N-1}{N} S^2 \quad \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij} - \bar{x}_i\}^2$$

وهو المطلوب إثباته .

ملحق رقم (٥ ـ ٥)

- إثبات أن :

$$V\{\overline{x}_{sy}\} = \frac{S^2}{n}\,\frac{N-1}{N}\,\{1+(n-1)r\}$$

حيث .

$$r = \frac{2}{\{n-1\}\{N-1\}S^2}\sum_{i=1}^{k}\sum_{j<j'}^{n}\{x_{ij}-\mu\}\{x_{ij'}-\mu\}$$

نعلم أن :

$$V\{\overline{x}_{sy}\} = E\{\overline{x}_i - E\{\overline{x}_i\}\}^2$$

$$= E\{\overline{x}_{sy}-\mu\}^2 = \frac{1}{k}\sum_{i=1}^{k}\{\overline{x}-\mu\}^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left|\frac{1}{n}\sum_{j=1}^{n}\{x_{ij}-\mu\}\right|^2$$

$$= \frac{1}{k}\,\frac{1}{n^2}\,\{\sum_{i=1}^{k}\sum_{j=1}^{n}\{x_{ij}-\mu\}^2 + 2\sum_{i=1}^{k}\sum_{j<j'}^{n}\{x_{ij}-\mu\}\{x_{ij'}-\mu\}$$

إن معامل الارتباط (r) بين كل زوج من الوحدات من العينة نفسها يساوى

$$r = \frac{E\{x_{ij}-\mu\}\{x_{ij'}-\mu\}}{E\{x_{ij}-\mu\}^2}$$

وعدد الأزواج المختلفة لوحدات المعاينة المنتظمة التي حجمها (n) وحدة هو

$$\binom{n}{2} = \frac{n!}{2!\{n-2\}!} = \frac{n\ (n-1)}{2}$$

زوجاً من الوحدات . وحيث لدينا (K) عينة ممكنة ، لذا فإن عدد الأزواج الممكنة هو $Kn\{n-1\}/2$ ويكون احتمال كل زوج $2/Kn(n-1)$ وبالتالي فإن :

$$E\{x_{ij}-\mu\}\{x_{ij'}-\mu\} = \frac{2}{kn\{n-1\}} \sum_{i=1}^{k} \sum_{j<j'}^{n} \{x_{ij}-\mu\}\{x_{ij'}-\mu\}$$

كما أن :

$$E\{x_{ij}-\mu\}^2 = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij}-\mu\}^2$$

$$= \frac{N-1}{N} \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{n} \{x_{ij}-\mu\}^2$$

$$= \frac{N-1}{N} S^2$$

ويكون معامل الارتباط {r} مساوياً :

$$r = \frac{2}{kn\{n-1\}} \sum_{i=1}^{k} \sum_{j<j'}^{n} \{x_{ij}-\mu\}\{x_{ij'}-\mu\} \frac{N}{\{N-1\} S^2}$$

ومنه نجد أن :

$$\sum_{i=1}^{k} \sum_{j<j'}^{n} \{x_{ij}-\mu\}\{x_{ij'}-\mu\} = \frac{n-1}{2} \frac{S^2\{N-1\} r}{1}$$

وبالتبديل فى $V\{\bar{x}_{sy}\}$ نجد أن :

$$V(\bar{x}_{sy}) = \frac{1}{k}\frac{1}{n^2}\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\{x_{ij}-\mu\}^2 + 2\,\frac{(n-1)}{2}\,\frac{(N-1)}{1}\,r\right\}$$

$$= \frac{1}{k}\frac{1}{n^2}\left[\sum_{i}^{k}\sum_{j}^{n}\{x_{ij}-\mu\}^2 + \{N-1\}S^2(n-1)r\}\right]$$

$$= \frac{1}{kn}\frac{1}{n}\{N-1\}S^2 + (\{N-1\}S^2(n-1)r\}$$

$$= \frac{S^2}{n}\frac{N-1}{N}(1+(n-1)r\}$$

وهو المطلوب إثباته :

– إثبات أن $\hat{V}\{\bar{x}_{ran}\}$ هو مقدّر غير متحيّز لـ $V\{\bar{x}_{sy}\}$ أى المطلوب إثبات أن

$$E\{\hat{V}\{\bar{x}_{ran}\} = V\{\bar{x}_{sy}\}$$

$$= \frac{N-1}{N}\frac{s^2}{n}\{1+\{n-1\}r\}$$

نعلم أن :

$$E\{\hat{V}\{\bar{x}_{ran}\} = E\left[\frac{N-n}{N}\frac{s^2}{n}\right]$$

$$= \frac{N-n}{N}\frac{1}{n}\,E\{s^2\}$$

$$= \frac{1}{n-1}\,E\left|\sum_{i=1}^{n}\{x_{ij}-\bar{x}\}^2\right| = \frac{1}{n-1}\left|E\sum_{j=1}^{n}x_{ij}^2 - n\,E\{\bar{x}\}^2\right|$$

إن

$$E\sum_{j=1}^{n}x_{ij}^2 - \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{n}x_{ij}$$

$$E\{\bar{x}^2\} = \bar{X}^2 + V\{\bar{x}_{sy}\}$$

إذن :

$$E\{s^2\} = \frac{1}{n-1}\left[\frac{1}{K}\sum_i^k \sum_j^n x_{ij}^2\right] - n\left[\overline{X}^2 + V(\overline{x}_{sy})\right]$$

$$= \frac{1}{n-1}\frac{1}{K}\left|\sum_i^k \sum_j^n x_{ij}^2\right| - kn\ \overline{X}^2 - nk\{\overline{x}_{sy}\}$$

$$= \frac{1}{n-1}\ \frac{1}{K}\ \left|(N-1)s^2 - N\frac{s^2}{n}\ \frac{N-1}{N}\{1+(n-1)r\right|$$

$$= \frac{1}{n-1}\ \frac{1}{K}\ (N-1)s^2\{1 - \frac{1}{n}(1+(n-1)r\}$$

$$= \frac{1}{n-1}\ \frac{1}{K}\ (N-1)s^2\left(\frac{n-1}{n} - \frac{n-1}{n}r\right)$$

$$= \frac{N-1}{n-1}\ \frac{1}{k}\ s^2\left(\frac{n-1}{n}\right)\{1-r\} = \frac{N-1}{N}\ s^2\ \{1-r\}$$

وبتبديل قيمة $E\{s^2\}$ بقيمتها نجد أن :

$$E\{\widehat{V}(x_{ran})\} = \frac{N-n}{N}\ \frac{1}{n}\ \frac{N-1}{N}\ s^2(1-r)$$

$$= \frac{N-1}{N}\ \frac{s^2}{n}\ \frac{(N-n)}{N}\ (1-r)$$

$$= \frac{N-1}{N}\ \frac{s^2}{n}\ \left(1-\frac{n}{N}\right)(1-r)$$

$$= \frac{N-1}{N}\ \frac{s^2}{n}\ \left[(1-r) - n\frac{(1-r)}{N}\right]$$

ونعلم أن : $r = \frac{-1}{N-1}$

ومن هذه العلاقة نجد أن : $N = 1 - \frac{1}{r} = \frac{r-1}{r}$

إذن

$$E\{\hat{v}(\overline{x}_{ran})\} = \frac{N-1}{N}\frac{s^2}{n}\left[1 - r - n\{1 - r\}\frac{r}{r-1}\right]$$

$$= \frac{N-1}{N}\frac{s^2}{n}\left\{1 - r + nr\frac{(r-1)}{r-1}\right\}$$

$$= \frac{N-1}{N}\frac{s^2}{n}\{1 + \{nr - r\}\}$$

$$= \frac{N-1}{N}\frac{s^2}{n}\{1 + \{n-1\}r\}$$

وهو المطلوب إثباته .

المراجع

باللغة العربية :

أحمد عباده سرحان : العينات ، مكتبة النهضة المصرية ، القاهرة ١٩٥٧م .

- بول . ج هويل : المبادئ الأولية في الإحصاء (ترجمة ومراجعة بدرية عبدالوهاب ، ومحمد الشربيني) ، دار جون وايلي وأبنائه ، نيويورك ١٩٨٤م .

- جلال مصطفى الصياد ومصطفى جلال مصطفى : مقدمة في طرق المعاينة الإحصائية ، مكتبة مصباح ، جدة ١٩٩٠م .

- حنان عيسى سلطان وغانم سعيد العبيدي : أساسيات البحث العلمي ، دار العلوم للطباعة والنشر ، الرياض ١٩٨٤م .

- خالد بالطيور: مقدمة في التحليل الإحصائي مع برنامج ساس ، مؤسسة جمال الجاسم للإلكترونيات ، الدمام ١٩٩٠م .

- نوقان عبيدان وعبدالرحمن عدس وكايد عبدالحق : البحث العلمي ، عمان ١٩٨٢ م .

محمد صبحي أبو صالح وعدنان محمد عوض : مقدمة في الإحصاء ، دار جون وايلي وأبنائه ، نيويورك ١٩٨٣م .

باللغة الإنجليزية :

- Aronson, M. & A. : "SAS System : A Programer's Guide", McGraw - Hill Inc., U.s., 1990.
- Cochran, W . : "Sampling Techniques", John Wiley & Sons, New York, 1977.
- Kish L. : "Survey Sampling", John Wiley & Sons, New York, 1965.
- Ryan & Others : Minitab, Duxbury Press, Posten, 1985.
- Thompson S. k : "Sampling", John Wiley. & Sons , New York 1992.
- SAS Institute : "SAS User's Guide {Statistics}", Ver. 5 Edition, Sas Institute Inc., 1985.
- SAS Institute : "SAS User's Guide (Basic)", Ver. 5 Edition, Sas Institute Inc., 1985.
- Scheaffer, : Mendenhall Ott : "Syrvey Sampling", Duxbury Prss, Massachusetts, 1979.
- Yates F : "Sampling Methods For Censuses And Surveys" : Charles Griffin & Co. Ltd. London, 1981.

×× المؤلف في سطور :

د. عبدالرزاق أمين مصطفى أبو شعر .

من مواليد دمشق ـ الجمهورية العربية السورية ، عام ١٩٤٣م .

المؤهل العلمي :

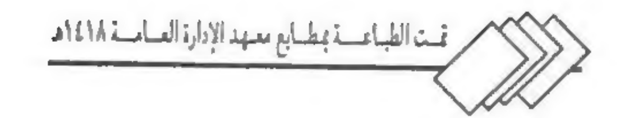
دكتوراه في الإحصاء من جامعة دمشق ، بالجمهورية العربية السورية ـ عام ١٩٨٨م .

خبرته العملية :

عمل في عدة وظائف في مجال الإحصاء في سورية والمملكة العربية السعودية ، وعمل عضواً بهيئة التدريب في معهد الإدارة العامة بالرياض .

الأنشطة العلمية :

- تنفيذ عدد من البحوث الميدانية و المكتبية .
- نشر عدداً من المقالات في مجلة (الإدارة العامة) ومجلة (جامعة دمشق) .
- تدريس عدد من مواد الإحصاء في معهد الإدارة العامة ، وجامعة حلب ، ومركز التدريب الإحصائي ، والمعهد التجاري بجامعة دمشق .



تمت الطباعة بمطابع معهد الإدارة العامة ١٤١٨هـ

٣٢ ريالاً

ردمك : ٩٩٦٠ - ١٤ - ٠٢١ - ٠

تمت الطباعة بمطابع معهد الإدارة العامة - ١٤١٨هـ